A NEW INFORMATION COCYCLE WITH SOME APPLICATIONS.

Roger Butler

Thesis submitted for the degree of Doctor of Philosophy.

Mathematics Institute

University of Warwick

July 1982.

To my parents and my friends.

TABLE OF CONTENTS.

Acknowledgements.

First and foremost, I would like to thank my supervisor Dr. Klaus Schmidt for his constant help, advice and encouragement. The generous friendship of both Klaus and his wife Annelise has helped to bring this thesis to fruition. For all these things and more, I am deeply grateful.

Special thanks are due to Professor Bill Parry, Dr. Peter Walters and many other friends for the interest they have shown in my work, and for discussing mathematics with me.

I am personally greatly indebted to my parents for their patient encouragement and support.

My sincere thanks to Peta McAllister for the excellent job she has made of typing this thesis.

It is a pleasure to thank both Professor T. Hamachi and Cambridge University Press for permission to include as an appendix the paper "On a Bernoulli shift with non-identical factor measures" as it appeared within the Journal of Ergodic Theory and Dynamical Systems, Volume 1.

Finally, I would like to thank the Science Research Council for the financial support they provided during the years 1977-1980.

Declaration.

The contents of Section 3 are contained in a joint paper ([B&S]) with Klaus Schmidt, and we follow very closely the exposition given there, which is mostly due to Klaus. The nature of the joint work is the following.

The first subsection of Section 3 contains notation, and background results which are in general well known, apart from Lemma 3.5 which is original and my own.

With the exceptions of Theorem 3.25 and Lemma 3.26, the results contained in the subsections entitled "A cohomology", and "The information cocycle" are due to me, but they were put into their present context by Klaus Schmidt. Theorem 3.25 and Lemma 3.26 should be viewed as joint work.

The problems and results discussed in the subsection called "Entropy and the information cocycle" arose from discussions with Bill Parry and Klaus Schmidt, the proofs however are due to Klaus.

In the final subsection, entitled "The information cocycle and the entropy of Anosov maps", the proofs provided are my own, and do not appear in this form within [B&S]. A closely related proof of Theorem 3.29 is however given there.

Summary.

The primary aim of this thesis is to present an information cocycle for groups of non-singular transformations.

In Section 1 we produce some coding results for Lipschitz partitions that are later used to associate canonically an information cocycle to Axiom A^* homeomorphisms.

As a specific example of the use of coding techniques, in Section 2 we prove that a Markov shift over finitely many states is isomorphic to its inverse, by a finitary isomorphism with finite expected code lengths. We also show that reversibility is a generic property.

In Section 3, an information cocycle is defined for groups of non-singular transformations. Some of its properties are then investigated, including its close relation to entropy, and an application is presented.

The information cocycle for single non-singular transformations is considered in Section 4, where we also study several conditions that imply that two σ-algebras are I-related.

In Section 5, the form of the information cocycle is calculated for generalised Markov shifts, and finite co-ordinate changes on shift spaces. Some invariants of the information cocycle-coboundary equation are also produced.

Section 6 contains two constructions for producing ergodic non-singular Bernoulli measures that are not equivalent to shift invariant probability measures.

Introduction.

Since reasonably comprehensive introductions will be provided at the beginning of each section, we shall give here only a brief overview of some of the connections between the various topics involved.

Of fundamental importance within this thesis, is the introduction in Section 3 of an information cocycle for groups of non-singular transformations. For measure preserving endomorphisms this information cocycle is shown to be intimately connected with entropy.

Using coding techniques produced in Section 1, we are able in Section 4 to canonically associate an information cocycle to Axiom A^* homeomorphisms of compact metric spaces. Coding results from Section 1 are also used in Section 3 to provide a simple proof of a well known formula for the entropy of an Anosov diffeomorphism.

The constructions of Section 6 are included to provide examples of ergodic transformations that do not admit an equivalent σ-finite invariant measure, and hence for which the extension of the classical information function to non-singular transformations is of particular interest.

In Sections 3 and 4, an information cocycle-coboundary equation appears, and within Section 5 some invariants of this equation are produced.

Section 2 is somewhat unrelated to the rest of the thesis, however we do introduce in it many of the definitions and notations used for Markov shifts. It also provides a specific example of the use of coding techniques.

SECTION 0.

BACKGROUND AND NOTATION.

In this section we introduce some of the definitions and notation that will be used throughout the rest of the thesis. A list of properties of the conditional information function is also included.

Throughout this thesis (except where indicated to the contrary), $(X,\mathcal{B},m)$ will denote a Lebesgue probability space, and all partitions will be assumed (or proved) to be measurable, in the sense of Rohlin [R1].

0.1. DEFINITIONS.

For α,β finite or countable ordered partitions of X with the same cardinality (where $\alpha = \{A_i\}$, $\beta = \{B_i\}$, $i \in I$ some finite or countable index set), we define a distance between α and β by

$$d_0(\alpha,\beta) := \sum_{i\in I} m(A_i \,\Delta\, B_i) \ ,$$

where Δ denotes the symmetric difference.

If α is a finite or countable partition, and $C \subset \mathcal{B}$ is a sub-σ-algebra, we define an asymmetric distance between α and C by

$$d(\alpha,C) := \inf\{d_0(\alpha,\beta);\ \beta \subset C,\ \mathrm{card}(\beta) = \mathrm{card}(\alpha)\} \ .$$

The above definition of d may be extended to a distance between two sub-σ-algebras as follows.

For sub-σ-algebras $A,C \subset \mathcal{B}$, we define

$$\begin{aligned} d(A,C) &:= \sup\{d(\alpha,C);\ \alpha \subset A\} \\ &= \sup_{\alpha\subset A}\ \inf_{\beta\subset C}\ \{d_0(\alpha,\beta);\ \mathrm{card}(\alpha) = \mathrm{card}(\beta)\} \ . \end{aligned}$$

We now list some properties of the distance d .

0.2. PROPERTIES OF d .

From the above definitions we see that for A_1, A_2, A_3 and C sub-σ-algebras of B , the distance d satisfies:

$$0 \le d(A,C) \le 2 , \tag{0.1}$$

$$d(A,C) = 0 \text{ if and only if } A \subset C . \tag{0.2}$$

As a consequence of the triangle inequality for symmetric differences

$$d(A_1,A_3) \le d(A_1,A_2) + d(A_2,A_3) , \tag{0.3}$$

and hence using property (0.2),

$$d(A_1,A_3) \le d(A_1,A_2) \text{ when } A_2 \subset A_3 , \tag{0.4}$$

$$d(A_1,A_3) \le d(A_2,A_3) \text{ when } A_1 \subset A_2 . \tag{0.5}$$

If A, C, A_n, $n = 1,2,\ldots$, are sub-σ-algebras of B and $A_n \uparrow A$, (i.e. $A_1 \subset A_2 \subset \ldots$ and $\bigcup_n A_n$ generates A) , then

$$d(A,C) = \lim_{n\to\infty} d(A_n,C) . \tag{0.6}$$

If further, α is a countable partition, then

$$d(\alpha,A) = \lim_{n\to\infty} d(\alpha,A_n) \ . \tag{0.7}$$

We also have for sub-σ-algebras $A,C \subset B$

$$d(A\vee C,C) = d(A,C), \tag{0.8}$$

and if T is a measure preserving endomorphism of (X,B,m) then

$$d(A,C) = d(T^{-1}A,T^{-1}C) \ . \tag{0.9}$$

One other useful property of d is that for A_1,A_2,C_1 and C_2 sub-σ-algebras of B,

$$d(A_1\vee A_2,C_1\vee C_2) \le d(A_1,C_1) + d(A_2,C_2) \ . \tag{0.10}$$

0.3. DEFINITION AND NOTATION.

Let us define a symmetric distance between sub-σ-algebras $A,C \subset B$ by setting

$$D(A,C) := \max\{d(A,C),d(C,A)\} \ ,$$

then D is a metric on the set of all sub-σ-algebras of B .

When no confusion can arise, we shall often use the 1-1 correspondence between (equivalence classes of) sub-σ-algebras and (equivalence classes of) measurable partitions (see [R1]) to identify notationally a measurable partition with its associated sub-σ-algebra. For situations where confusion could occur, if A is a sub-σ-algebra, we use $\hat{A}$ to denote the associated measurable partition; conversely if α is a measurable partition we denote by $\hat{\alpha}$ the corresponding sub-σ-algebra.

With a fixed endomorphism T of (X,B,m) in mind, and a partition α, we shall write α_{-k}^{ℓ} for $\bigvee_{i=-k}^{\ell} T^{-i}\alpha$, α^{-} for $\bigvee_{i=0}^{\infty} T^{-i}\alpha$, and α_T for $\bigvee_{i=-\infty}^{\infty} T^{-i}\alpha$.

If more than one measure is in use, we shall use a subscript to indicate the measure with respect to which we are working, for instance $d_m(\cdot,\cdot)$ denotes the distance $d(\cdot,\cdot)$ taken with respect to the measure m.

Further details of the above definitions and properties may be found in [P2], [P3], [Bu], and [T2].

0.4. DEFINITION.

If X is also a metric space, then for a measurable partition α, we define

$$\operatorname{diam}(\alpha) := \inf_{\alpha'} \{\sup_{A'\in\alpha'} \{\operatorname{diam}(A')\};\ \alpha' \text{ is a partition with } D(\alpha',\alpha) = 0\}.$$

We now define the concept of conditional expectation, which will be extensively used in Sections 3,4 and 5.

0.5. <u>DEFINITION</u>.

For a sub-σ-algebra $A \subset B$ and an integrable function f, we denote by $E_m(f|A)$ the *conditional expectation* of f with respect to the σ-algebra A, and the measure m. The conditional expectation $E_m(f|A)$ is uniquely defined (modulo A-measurable sets of measure zero), by the conditions

a) $E_m(f|A)$ is A-measurable.

b) $\int_A E_m(f|A)dm = \int_A fdm$,

for all sets $A \in A$.

An important and fundamental property of conditional expectation is described by the following theorem.

0.6. <u>THE INCREASING MARTINGALE THEOREM</u>.

Let A_n be an increasing sequence of sub-σ-algebras of B, and let A be the sub-σ-algebra that $\bigcup_n A_n$ generates, then for every integrable function f,

$$E_m(f|A_n) \to E_m(f|A) \text{ m-a.e. and in } L^1(X,B,m) .$$

For a proof of this result see [D] or [N].

Another concept extensively used in this thesis, is the information function. This will be defined in Section 3.

0.7. PROPERTIES OF THE INFORMATION FUNCTION.

For convenience of reference we list here some of the basic properties of information for partitions (and hence by extension sub-σ-algebras of $\mathcal{B}$). If N denotes the trivial σ-algebra that consists of sets of measure one and zero, then we write $I(A|N)$ as $I(A)$.

Let α,β,γ be countable partitions, then

(i) $I(\alpha|\gamma) \geq 0$, with equality if and only if $\alpha \subset \gamma$. (0.11)

(ii) $I(\alpha) \geq I(\beta)$ when $\alpha \supset \beta$. (0.12)

(iii) $I(\alpha|\gamma) \geq I(\beta|\gamma)$ when $\alpha \supset \beta$. (0.13)

(iv) $I(\alpha\vee\beta) = I(\alpha) + I(\beta|\alpha)$. (0.14)

(v) $I(\alpha\vee\beta|\gamma) = I(\alpha|\gamma) + I(\beta|\alpha\vee\gamma)$. (0.15)

(vi) $I(\alpha\vee\beta|\gamma) = I(\alpha|\gamma)$ when $\beta \subset \gamma$. (0.16)

(vii) $I(\alpha|\gamma)\circ T = I(T^{-1}\alpha|T^{-1}\gamma)$, (0.17)

for T a measure preserving endomorphism of $(X,\mathcal{B},m)$.

For further details and proofs of the above properties, we refer the reader to [B], [P1], [P4], or [Pi] . As general references for work related to the contents of this thesis, see for example [M&E], [P&T2] and [W1].

SECTION 1.

CODING OF PARTITIONS.

In [Bo2], R. Bowen used coding techniques to prove that small smooth partitions of C^2 Anosov diffeomorphisms are weak Bernoulli. To do this he introduced the idea of bounded coding between two finite partitions (with respect to a transformation), and showed that for C^2 Anosov diffeomorphisms, small smooth finite partitions boundedly code each other. Subsequently, in [P3], W. Parry used the ideas of bounded codes and ε-bounded codes to produce the same information cocycle-coboundary equation as appeared in [F&P]. Using invariants of this equation he was then able to distinguish between particular transformations with respect to various restrictive types of isomorphisms.

In this section we extend R. Bowen's result concerning bounded coding, to the case of Lipschitz partitions of Axiom A^* homeomorphisms on compact metric spaces. We also introduce a particular type of endomorphism of compact metric spaces, called Lipschitz expanding, and prove a similar result for these maps. Examples of Lipschitz expanding maps are C^1-expanding maps on compact, connected, C^∞ Riemannian manifolds.

The main use of the coding results presented in this section occurs within Section 4, where we use them to associate an information cocycle to Axiom A^* homeomorphisms (see Corollary 4.12), in a canonical manner. Specifically, we use the information cocycle corresponding to the past σ-algebra of a small Lipschitz partition, and then utilise the results of Sections 3 and 4 to show that such an association is canonical.

In Section 3 similar coding techniques are used to prove a well known formula for the entropy of a C^2 Anosov diffeomorphism.

1.1. DEFINITION.

If α,β are two partitions of a Lebesgue probability space $(X,\mathcal{B},m)$, then β *ε-boundedly codes* α, with respect to a non-singular endomorphism $T:X \to X$ and the measure m, if there exists some positive integer $k = k(\varepsilon)$ such that

$$d(\alpha_k^{n+k}, \beta_0^{n+2k}) \le \varepsilon \tag{1.1}$$

for all $n = 0,1,2,\ldots$.

1.2. Remark.

In the case of a measure preserving automorphism T, this definition is equivalent to the standard one (β ε-boundedly codes α if there exists a positive integer $k = k(\varepsilon)$ such that $d(\alpha_0^n, \beta_{-k}^{n+k}) \le \varepsilon$ for all $n = 0,1,\ldots$) by virtue of property (0.9) of $d(\cdot,\cdot)$.

1.3. DEFINITION.

We say that β *boundedly codes* α, if for every $\varepsilon > 0$, β ε-boundedly codes α.

1.4. Remark.

For a fixed automorphism T, probability measures μ,ν satisfying $\mu \ll \nu$, and partitions α,β ; β boundedly codes α with respect to T and ν implies that β boundedly codes α with respect to T and μ. This is a consequence of the fact that for such measures μ and ν ; for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any measurable set B, $\nu(B) < \delta$ implies that $\mu(B) < \varepsilon$ (see [H2; p.125, Th.B]).

1.5. LEMMA.

Let T be a measure preserving endomorphism of the Lebesgue probability space $(X,\mathcal{B},m)$, and let α,β be two partitions that satisfy (with respect to T and m),

$$\sum_{i=0}^{\infty} d(T^{-i}\alpha,\beta_0^{2i}) < \infty , \tag{1.2}$$

then β boundedly codes α with respect to T and m.

Proof.

Given $\varepsilon > 0$, choose k such that

$$\sum_{i=k}^{\infty} d(T^{-i}\alpha,\beta_0^{2i}) \le \varepsilon/2 ,$$

then

$$\begin{aligned} d(\alpha_k^{n+k},\beta_0^{n+2k}) &\le \sum_{i=k}^{n+k} d(T^{-i}\alpha,\beta_0^{n+2k}) \\ &\le 2\sum_{i=k}^{\infty} d(T^{-i}\alpha,\beta_0^{2i}) \\ &\le \varepsilon , \text{ for all } n = 0,1,2,\dots . \end{aligned}$$

Since the choice of ε was arbitrary, by Definitions 1.1 and 1.3, the result is proved. □

1.6. Remark.

In the case of T a measure preserving automorphism, condition (1.2) can be rewritten as

$$\sum_{i=0}^{\infty} d(\alpha,\beta_{-i}^{i}) < \infty \quad . \tag{1.3}$$

1.7. NOTATION.

Let f be a homeomorphism of a metric space (X,ρ) , then we define for $x \in X$ and $\varepsilon > 0$,

$$W_{\varepsilon}^{s}(x) := \{y \in X; \rho(f^n(x),f^n(y)) \le \varepsilon \text{ for all } n \ge 0\} \quad ,$$

$$W_{\varepsilon}^{u}(x) := \{y \in X; \rho(f^n(x),f^n(y)) \le \varepsilon \text{ for all } n \le 0\} \quad ,$$

$$W^{s}(x) := \{y \in X; \rho(f^n(x),f^n(y)) \to 0 \text{ as } n \to +\infty \} \ ,$$

$$W^{u}(x) := \{y \in X; \rho(f^n(x),f^n(y)) \to 0 \text{ as } n \to -\infty \} \ .$$

The following class of transformations was introduced and studied by R. Bowen in [Bo1]. For such transformations we shall later prove a coding result.

1.8. DEFINITION.

A homeomorphism f of a compact metric space (X,ρ) is said to *satisfy Axiom* A^* precisely when all of the following conditions are satisfied:

A1. The periodic points of f are dense in X .

A2. For each $\varepsilon > 0$, there is a $\delta = \delta(\varepsilon) > 0$ such that $W_{\varepsilon}^{s}(x) \cap W_{\varepsilon}^{u}(y) \ne \emptyset$ whenever $\rho(x,y) \le \delta$.

A3. There are $\varepsilon^* > 0$, $0 < \lambda < 1$ and $c \ge 1$ such that for all $n \ge 0$,

$$\rho(f^n(x),f^n(y)) \le c\lambda^n\rho(x,y) \quad \text{if} \quad y \in W^s_{\varepsilon^*}(x) \text{ , and}$$

$$\rho(f^{-n}(x)\, ,\, f^{-n}(y)) \le c\lambda^n\rho(x,y) \quad \text{if} \quad y \in W^u_{\varepsilon^*}(x) \text{ .}$$

1.9. NOTATION.

Let f be a homeomorphism of a compact metric space (X,ρ), satisfying Axiom A^*. Choose ε, $0 < \varepsilon < \varepsilon^*/2$ (see A3) so that if $\rho(x,y) < \varepsilon$, then both $\rho(f(x),f(y)) < \varepsilon^*/2$ and $\rho(f^{-1}(x),f^{-1}(y)) < \varepsilon^*/2$. Let $\delta = \delta(\varepsilon)$ be as in A2, then if $\rho(x,y) < \delta$,

$$\begin{aligned} W^s_\varepsilon(x) \cap W^u_\varepsilon(y) &= W^s_{\varepsilon^*/2}(x) \cap W^u_{\varepsilon^*/2}(y) \\ &= 1\text{-point.} \end{aligned}$$

This is because ε^* is an expansive constant for f (see [Bo3]). If $x,y \in X$ satisfy $\rho(x,y) < \delta$, then we define a local product structure on X by setting

$$[x,y] := W^s_\varepsilon(x) \cap W^u_\varepsilon(y) \text{ .}$$

The following result is well-known for Axiom A diffeomorphisms. For reasons of completeness we include a straightforward proof for the case of Axiom A^* homeomorphisms. The essence of the proof is due to D. Ruelle (see [Ru; Ch.7]).

1.10 LEMMA.

Let f be a homeomorphism of a compact metric space (X,ρ), satisfying Axiom A^*. There exist constants $\delta > 0$, $C > 0$, and

$0 < \lambda < 1$ such that if $\rho(f^i(x),f^i(y)) < \delta$ for all $i \in [-n,\dots,n]$, then $\rho(x,y) < C\lambda^n$.

Proof.

Choose ε and δ corresponding to f, as in Notation 1.9, and let $c \geq 1$, and $0 < \lambda < 1$ be chosen as in A3 of Definition 1.8.

Suppose $x,y \in X$ satisfy $\rho(f^i(x),f^i(y)) < \delta$ for all $i \in [-n,\dots,n]$, then

$$\rho(f(x),f([x,y])) < \varepsilon^*/2\ ,\quad \text{because}\quad \rho(x,[x,y]) < \varepsilon\ ,$$

and

$$\rho(f([x,y]),f(y)) < \varepsilon^*/2\ ,\quad \text{because}\quad \rho([x,y],y) < \varepsilon\ .$$

Together these inequalities imply that

$$\begin{aligned} f([x,y]) &\in W^s_{\varepsilon^*/2}(f(x)) \cap W^u_{\varepsilon^*/2}(f(y)) \\ &= W^s_\varepsilon(f(x)) \cap W^u_\varepsilon(f(y))\ ; \quad \text{as}\quad \rho(f(x),f(y)) < \delta \\ &= [f(x),f(y)]\ . \end{aligned}$$

Similarly, for all $i \in [-n,\dots,n]$,

$$f^i([x,y]) = [f^i(x),f^i(y)] = W^s_\varepsilon(f^i(x)) \cap W^u_\varepsilon(f^i(y))\ .$$

Since $f^{-n}([x,y]) \in W^s_\varepsilon(f^{-n}(x)) \subset W^s_{\varepsilon^*}(f^{-n}(x))$, condition A3 (of Definition 1.8) implies that

$$\rho(x,[x,y]) \leq c\lambda^n\varepsilon\ .$$

Also, as $f^n([x,y]) \in W^u_\varepsilon(f^n(y)) \subset W^u_{\varepsilon^*}(f^n(y))$,

$$\rho([x,y],y) \leq c\lambda^n\varepsilon \quad ,$$

and hence

$$\rho(x,y) \leq 2\varepsilon c\lambda^n \; .$$

Because both ε and c are independent of x and y , by defining $C: = 3\varepsilon.c$, the result is proved. □

1.11. DEFINITION.

For α a partition of a metric space, let us denote the set of all points within a distance ε of the boundary of some element of α , by $\partial_\varepsilon\alpha$.

1.12. DEFINITION.

Let (X,ρ) be a compact metric space, and m be a Borel probability measure on X . If α is a measurable partition of X , then it is called a *Lipschitz partition* (with respect to the metric ρ and the measure m) precisely when there exists a constant $c = c(\alpha)$ such that for all $\varepsilon > 0$, $m(\partial_\varepsilon\alpha) \leq c.\varepsilon$.

For Lipschitz partitions and Axiom A^* homeomorphisms we have the following coding result. Note that for C^2 Anosov diffeomorphisms a similar result was proved by R. Bowen in [Bo2].

1.13. PROPOSITION.

Let (X,ρ) be a compact metric space, and m be a Borel probability measure on X. If $f:X \to X$ is a homeomorphism satisfying Axiom A^* and preserving the measure m, then with respect to f and m, Lipschitz partitions are boundedly coded by small partitions.

Proof.

Let α be a fixed Lipschitz partition, and let β be any (measurable) partition of X with $\operatorname{diam}(\beta) < \delta$, where δ is the constant appearing in Lemma 1.10 and Notation 1.9. From Lemma 1.10, there exists a constant C such that for all $i \in \mathbb{N}$, $\operatorname{diam}(\beta_{-i}^{i}) \le C\lambda^i$. Since α is Lipschitz this implies that

$$\sum_{i=0}^{\infty} d(\alpha,\beta_{-i}^{i}) \le \sum_{i=0}^{\infty} 2.c(\alpha).C\lambda^i < \infty , \tag{1.4}$$

where $c(\alpha)$ is the constant from Definition 1.12 corresponding to α, and $d(\alpha,\beta_{-i}^{i}) \le c(\alpha).\operatorname{diam}(\beta_{-i}^{i})$ for each i because the total measure of all elements of β_{-i}^{i} intersecting more than one element of α is less than $c(\alpha).\operatorname{diam}(\beta_{-i}^{i})$. Applying Lemma 1.5 together with Remark 1.6 completes the proof.

□

1.14. <u>COROLLARY (of the proof)</u>.

Let f be as above, and let α be any Lipschitz partition of X. There exist constants $\delta > 0$ (independent of α), $c > 0$ and $0 < \lambda < 1$ such that if β is a partition with $\operatorname{diam}(\beta) < \delta$, then

$$d(\alpha_0^n,\beta_{-k}^{n+k}) \le c\lambda^k ,$$

for all $n,k = 0,1,2,\ldots$.

Proof.

Let δ be as in Proposition 1.13, and let α,β be as in the above hypotheses, then from equation (1.4) and property (0.10) of $d(\cdot,\cdot)$ we have

$$d(\alpha_0^n,\beta_{-k}^{n+k}) \le \sum_{i=0}^{n} d(\alpha,\beta_{-k-i}^{k+n-i})$$

$$\le 2\sum_{i=k}^{\infty} d(\alpha,\beta_{-i}^{i})$$

$$\le 4\sum_{i=k}^{\infty} c(\alpha).C\lambda^i \quad ,$$

for all $n,k = 0,1,2,\ldots$.

By defining

$$c: = 4\sum_{i=0}^{\infty} c(\alpha).C\lambda^i \quad ,$$

the result is proved. Note that δ is independent of α .

□

The following is an immediate corollary of Proposition 1.13.

1.15. COROLLARY.

Let f and m be as above, then with respect to f and m , small Lipschitz partitions boundedly code each other.

We now consider a particular class of endomorphisms that possess a local expansion property. The existence of invariant measures and equilibrium states for these and other closely related maps was studied by P. Walters in [W2].

1.16. DEFINITION.

Let (X,ρ) be a compact metric space. A map $f:X \to X$ is called *Lipschitz expanding* precisely when it is a continuous surjection satisfying:

a) There exist constants $b > 0$, $a > 0$ and $\lambda > 1$ such that for all $x,y \in X$, if for some $m \in \mathbb{N}\setminus\{0\}$

$$\rho(f^i(x),f^i(y)) < b \quad \text{for all} \quad i = 0,1,\dots,m-1 ,$$

then

$$\rho(f^m(x),f^m(y)) \geq a\lambda^m \rho(x,y) . \tag{1.5}$$

1.17. Remark.

Note that the constant $b > 0$ appearing in the above definition is small enough to guarantee that if $0 < \rho(x,y) < b$ then $f(x) \neq f(y)$, i.e. locally only expansion occurs. In particular, since the maximum cardinality of a b-separated set in a compact metric space is finite (see for example [W1; Th.6.4.]), this implies that f is bounded-to-one.

The above definition of a Lipschitz expanding map is valid for any equivalent metric, however the constants that appear may differ.

The main result of this section can now be stated as follows:

1.18. THEOREM.

Let (X,ρ) be a compact metric space and m be a Borel probability measure on X . If $f:X \to X$ is a Lipschitz expanding map preserving the measure m , then with respect to f and m , Lipschitz partitions are boundedly coded by small partitions.

Proof.

Let $b > 0$, $a > 0$, and $\lambda > 1$ (respectively) be constants with respect to which f is Lipschitz expanding (see Definition 1.16).

We shall first prove that there exist real numbers $\delta > 0$, $C < 1$ and an integer r , such that for partitions β satisfying $\mathrm{diam}(\beta) < \delta$, the following inequality holds:

$$\mathrm{diam}(\beta \vee f^{-r}\beta \vee \ldots \vee f^{-nr}\beta) \le C^n \mathrm{diam}(\beta) , \tag{1.6}$$

for all $n = 1,2,\ldots$.

The idea of the proof is to guarantee that disjoint "components" of pre-images are sufficiently separated, despite the local expansion property of f . We achieve this as follows.

Choose and fix an integer r such that $a\lambda^r > 1$, and (using the continuity of f) a positive constant $b' < b$ such that for any $x,y \in X$, if $\rho(x,y) < b'$, then $\rho(f^i(x),f^i(y)) < b$, for all $i = 0,1,\ldots,r$.

Since f is a Lipschitz expanding map, for any fixed $x \in X$, the set $E_x := \{y \in X; f^r(y) = x\}$ is finite. Let

$$t_x := \min\{\rho(y,z); y,z \in E_x \text{ and } y \ne z\}$$

and

$$s_x := \frac{1}{3} \min\{t_x, b'\} .$$

Using inequality (1.5) and the choice of b' we choose for each $x \in X$ a real number $\varepsilon_x > 0$ such that

$$f^{-r}(B(x;\varepsilon_x)) \subset \bigcup_{y \in E_x} B(y;s_x) ,$$

where for $z \in X$ and $\varepsilon > 0$, $B(z;\varepsilon)$ denotes the open ball of radius ε and centre z. The collection of all such sets $\{B(x;\varepsilon_x)\}_{x\in X}$ covers X, and using compactness we choose a finite subcover, denoted by $\mathcal{U} = \{U_i\}_{i\in I}$. From the construction of $\mathcal{U}$, to each set U_i, $i \in I$, there corresponds a point $x_i \in X$ such that $U_i = B(x_i;\varepsilon_{x_i})$. With respect to these points we define

$$d_i := \min\{\rho(B(y;s_{x_i}),B(z;s_{x_i}));y,z\in E_{x_i},y \neq z\} ,$$

and

$$d := \min\{d_i;i\in I\} .$$

Note that for any point $z \in f^{-r}(U_i)$, $i \in I$, the point z is contained in a unique set $B(y;s_{x_i})$, for some $y \in E_{x_i}$, and we shall refer to this set as the component of $f^{-r}(U_i)$ containing z.

We now define the number δ by setting

$$\delta := \frac{1}{2}\min\{L(\mathcal{U}),d,b'\} ,$$

where $L(\mathcal{U})$ denotes a Lebesgue number of the open cover $\mathcal{U}$ (see [Du; Ch.XI, Th. 4.5]).

If β is a partition of X with $\mathrm{diam}(\beta) < \delta$, then (neglecting sets of measure zero), every set $B \in \beta$ is contained within some $U_i \in \mathcal{U}$, and the components of $f^{-r}(U_i)$ (and hence of $f^{-r}(B)$) are all separated by distances of at least δ. The properties of the constants δ and b' now guarantee that for any two points $x,y \in X$ in the same component of some set $f^{-r}(B)$, $B \in \beta$,

$$\rho(f^i(x),f^i(y)) < b \quad \text{for all} \quad i = 0,1,\dots,r \ .$$

As f is a Lipschitz expanding map this implies that for any set $B \in \beta$, and any component B' of $f^{-r}(B)$,

$$\mathrm{diam}(B') \le a^{-1}\lambda^{-r-1}.\mathrm{diam}(B) \ ,$$

and hence

$$\mathrm{diam}(\beta \vee f^{-r}\beta) \le a^{-1}\lambda^{-r-1}.\mathrm{diam}(\beta) \ .$$

Since the above argument is valid for any partition with diameter less than δ, if we define $C{:}= a^{-1}\lambda^{-r-1}$, then by applying a straightforward induction argument we have proved that inequality (1.6) holds.

In the following, β will denote a fixed (but arbitrary) partition satisfying $\mathrm{diam}(\beta) < \delta$.

Let α be a Lipschitz partition of X, and let $d = d(\alpha)$ be the constant associated to it in Definition 1.12, i.e. for all $\varepsilon > 0$, $m(\partial_\varepsilon \alpha) \le d.\varepsilon$. Since for any partition γ the total measure of all elements of γ intersecting more than one element of α is less than $d.\mathrm{diam}(\gamma)$; if $\mathrm{diam}(\gamma) < \zeta$ then $d(\alpha,\gamma) < 2d\zeta$. Thus using inequality (1.6),

$$\begin{aligned} d(\alpha,\beta_0^i) &\le 2d.C^{[i/r]}.\mathrm{diam}(\beta) \\ &< 2d.\delta.C^{[i/r]} \ , \end{aligned} \tag{1.7}$$

for all $i = 0,1,2,\dots$, where $[x]$ denotes the integer part of x.

The properties of $d(\cdot,\cdot)$; inequality (0.4) and equation (0.9), together with inequality (1.7) therefore imply that

$$\sum_{i=0}^{\infty} d(T^{-i}\alpha, \beta_0^{2i}) \le \sum_{i=0}^{\infty} d(\alpha, \beta_0^{i})$$

$$\le \sum_{i=0}^{\infty} 2d.\delta.C^{[i/r]}$$

$$< \infty \; , \tag{1.8}$$

since $C < 1$ and r is a fixed integer. By Lemma 1.5, inequality (1.8) implies that β boundedly codes α .

The choice of the Lipschitz partition α was arbitrary however, and β was only required to satisfy $\mathrm{diam}(\beta) < \delta$, where δ is independent of α . We have therefore proved that Lipschitz partitions are boundedly coded by small partitions.

□

The following corollaries are analogous to Corollaries 1.14 and 1.15, and their straightforward proofs are omitted.

1.19. COROLLARY.

Let f be as above, and let α be any Lipschitz partition of X . There exist constants $\delta > 0$ (independent of α) , $c > 0$ and $0 < L < 1$ such that if β is a partition with $\mathrm{diam}(\beta) < \delta$, then

$$d(\alpha_0^n, \beta_0^{n+k}) \le cL^k \quad ,$$

for all $n,k = 0,1,2,\ldots$.

1.20. COROLLARY.

Let f and m be as above, then with respect to f and m, small Lipschitz partitions boundedly code each other.

We now describe a class of finite partitions that are Lipschitz, and using these we prove a coding result for a particular class of Lipschitz expanding maps that have already been extensively studied (see for example [S], [K&S], [W2] and [Sh]).

1.21. DEFINITION.

A finite partition of a differentiable manifold M is *smooth* if the boundary of each set in the partition is a compact, piecewise-C^1, submanifold of M.

1.22. PROPOSITION.

Let α be a finite smooth partition of a Riemannian manifold M, then α is a Lipschitz partition with respect to any probability measure m defined by a continuous volume form on M.

Proof.

Since the partition α is smooth, and the manifold M possesses a Riemannian metric, we may cover the combined boundary of all the elements of α with a finite number of differentiable charts (U_i, ϕ_i) that satisfy:

(i) For each i, the image under ϕ_i of the boundary in U_i is either a hyperplane or 2 (or more) half-planes joined along a common boundary.

(ii) If M is n-dimensional, and we define on each set $\phi_i(U_i) \subset \mathbb{R}^n$ the measure μ_i equal to Lebesgue measure restricted to $\phi_i(U_i)$, then the Radon-Nikodym derivative $dm/d(\mu_i \circ \phi_i)$ is bounded above m-a.e. on U_i, by a constant B_i (say), for each i.

As each map ϕ_i is differentiable, there also exist constants $C_i > 0$ (for each i) such that if $(U_i \cap \partial_\varepsilon \alpha)$ represents that part of $\partial_\varepsilon \alpha$ that lies in U_i, then for all $\varepsilon > 0$

$$\mu_i(\phi_i(U_i \cap \partial_\varepsilon \alpha)) \leq C_i.\varepsilon \quad . \tag{1.9}$$

Combining condition (ii) and inequality (1.9) we have that for all $\varepsilon > 0$

$$m(U_i \cap \partial_\varepsilon \alpha) \leq B_i C_i.\varepsilon \quad .$$

Let $D := \max_i (C_i B_i)$, let N be the number of charts (U_i, ϕ_i), and let $c := ND$, then for all $\varepsilon > 0$,

$$m(\partial_\varepsilon \alpha) \leq c.\varepsilon \quad ,$$

and thus α is a Lipschitz partition.

□

1.23. NOTATION.

Let M be a compact, connected, C^∞ manifold without boundary, equipped with a Riemannian metric $\|\cdot\|$, and denote by TM the tangent bundle of M.

Following the notation used in [S], a map is of class C^n_- if it is of class C^{n-1} and its $(n-1)^{st}$ derivative is Lipschitz. A map is of class C^n_+ if it is of class C^n and its Jacobian is of class C^n_- .

1.24. DEFINITION.

A map $\phi:M \to M$ of class C^1 is called *expanding* if it is surjective, and there exist real numbers $\lambda > 1$, $a > 0$ such that

$$\|(d\phi^n)v\| \geq a\lambda^n \|v\| \quad ,$$

for all $v \in TM$ and $n = 1,2,\ldots$.

1.25. PROPOSITION.

Let $\phi:M \to M$ be an expanding map, then with respect to the metric ρ on M induced by $\|\cdot\|$, ϕ is Lipschitz expanding.

Proof.

Let $a' > 0$ and $\lambda' > 1$ be the constants from Definition 1.24 corresponding to ϕ . We first note that the Riemannian metric $\|\cdot\|$ can be adapted to provide an equivalent C^∞ Riemannian metric $\|\cdot\|_1$ and a constant $\lambda > 1$ such that

$$\|(d\phi)v\|_1 \geq \lambda \|v\|_1 \; ,$$

for all $v \in TM$. (For details, see for example [K&S;Lemma 1].) Denote by ρ and ρ_1 the metrics on M induced by $\|\cdot\|$ and $\|\cdot\|_1$ respectively.

Since ϕ is an expanding map it is both an N-fold covering map, for some $1 < N < \infty$, and a local diffeomorphism. As a consequence of these properties, there exists a finite open cover $V = \{V_j\}_{j\in J}$ of M such that (see [K&S; Lemma 4]):

a) The map ϕ is a diffeomorphism of each (disjoint) component of $\phi^{-1}(V_j)$ onto V_j, for all $j \in J$.

b) Corresponding to each pair of points $x,y \in V_j$, there exists a regular curve joining x and y that is contained within V_j, and the length of which (using $\|\cdot\|_1$) is equal to $\rho_1(x,y)$, for all $j \in J$.

Let $L(V)$ be a Lebesgue number of V, then using $L(V)$ in the role of b, and ϕ^{-1} in the role of f^{-r}, we can now use an argument analogous to part of that used in the proof of Theorem 1.18, to obtain a constant $b_1 > 0$ such that whenever $x,y \in X$ satisfy $\rho_1(x,y) < b_1$, then $\rho_1(\phi(x),\phi(y)) \geq \lambda\rho_1(x,y)$. Proceeding inductively: for any $x,y \in X$, if

$$\rho_1(\phi^i(x),\phi^i(y)) < b_1 \quad \text{for all } i = 0,1,\ldots,m-1 ; \text{ some } m \in \mathbb{N}\backslash\{0\}$$

then

$$\rho_1(\phi^m(x),\phi^m(y)) \geq \lambda^m\rho_1(x,y) .$$

The equivalence of the metrics ρ and ρ_1 therefore implies the existence of constants $b > 0$, $a > 0$ and $\lambda > 1$ (the same λ), with respect to which ϕ is Lipschitz expanding. $\square$

The following theorem is closely related to Theorem 1.18.

1.26. <u>THEOREM</u>.

Let $\phi: M \to M$ be a C^1_+ expanding map, and let m be a probability measure absolutely continuous with respect to the Lebesgue measure (class) of M, then with respect to ϕ and m, smooth finite partitions are boundedly coded by small partitions.

Proof.

By Remark 1.4, it is sufficient to prove the result for some probability measure μ, with respect to which m is absolutely continuous. Choose a probability measure μ defined by a volume element of class C^0_+ on M, then by assumption, $m \ll \mu$. We now apply a result of R. Sacksteder [S;Th.5.1] which states that for ϕ and μ as above, there is a volume element of class C^0_+ that defines a ϕ-invariant probability measure ν satisfying :

$$\theta^{-1}\mu(A) \leq \nu(A) \leq \theta\mu(A) , \tag{1.10}$$

for every measurable set A, where θ is a positive constant independent of A.

Inequality (1.10) implies that the measures μ and ν are equivalent, and hence we can use Proposition 1.25, together with Theorem 1.18 (with respect to the invariant measure ν) and Remark 1.4 to complete the proof.

$\square$

We conclude this section by describing a well-known class of examples of Lipschitz expanding maps.

1.27. Example.

Let f be a piecewise-C^1, continuous endomorphism of the unit circle $\mathbb{T}^1$. If there exists a constant $\lambda > 1$ such that (where defined) the derivative of f is greater than λ , then f is a Lipschitz expanding map on $\mathbb{T}^1$. When f is also piecewise-C^2 , a result due to A. Lasota and J. Yorke (see [L&Y]) shows that f possesses an invariant probability measure absolutely continuous with respect to Lebesgue measure.

Similar examples can be constructed on higher dimensional tori.

SECTION 2.

REVERSIBILITY OF TRANSFORMATIONS.

In [H&N], P. Halmos and J. von Neumann proved (among other things) that any ergodic measure preserving transformation with discrete spectrum is isomorphic to its own inverse. They also conjectured (see [H&N; page 348]) that the same is true for arbitrary measure preserving transformations. As a counterexample to this conjecture, H. Anzai [A] produced a measure preserving automorphism on the 2-torus that is not conjugate to its inverse.

If we define an automorphism to be reversible precisely when it is isomorphic to its inverse, then with the above results in mind it is natural to ask: what types of transformations are reversible? Another interesting question is: how restrictive can we be about choosing the reversing map?

As an example of a transformation that is reversible, and for which the reversing map has particularly nice properties, we prove in this section that finite state Markov shifts are reversible, and we produce a reversing map for them that is finitary and has finite expected coding lengths. That this result should be true was suggested by W. Parry.

Having proved the above result, in the second subsection of this section we go off at somewhat of a tangent and partially answer the related question: of what topological category is the set of reversible transformations? We prove that within the classes of non-singular and measure preserving automorphisms (on a non-atomic Lebesgue probability space) equipped with the uniform topology, there exists a dense G_δ set of transformations that are reversible. To contrast with this result, we then use H. Anzai's example to prove that (within the same settings) the set of irreversible transformations is also dense.

2.1. DEFINITIONS.

Let A be a finite or countable set, and define the doubly infinite product space

$$X := \prod_{n=-\infty}^{\infty} A_n \,, \quad \text{where } A_n = A \text{ for all } n \in \mathbb{Z} \,.$$

Upon the space X we define the (left-) shift transformation S by setting

$$(Sx)_n := x_{n+1} \,, \quad \text{for all } n \in \mathbb{Z} \text{ and } x = (x_n) \in X \,.$$

Depending upon whether A has finite or infinite cardinality, we shall identify it with (as appropriate) either $\{0,1,\ldots,k-1\}$ for $k = \mathrm{card}(A)$, or $\{0,1,2,\ldots\}$. The set A will be referred to as the *alphabet*, and sets of the form

$$[i_0 \ldots i_\ell]^j := \{x = (x_n) \in X;\ x_j = i_0, \ldots, x_{j+\ell} = i_\ell\} \,,$$

where $i_0, \ldots, i_\ell \in A$, are called *cylinders*. By $\mathcal{B}$ we denote the σ-algebra generated by all cylinder sets.

If $P = (P(i,j); i,j \in A)$ is a matrix on A, then it is called *stochastic* if for all $i \in A$, $P(i,\cdot)$ is a probability on A. The matrix P is called *irreducible* when for each $i,j \in A$ there exists an integer n such that $P^n(i,j) > 0$.

It is well known (see, for example [F]) that for any $k \times k$ irreducible stochastic matrix P, there exists a unique strictly positive probability vector $p = (p(0),\ldots,p(k-1))$ such that $p.P = p$. Define for each cylinder

$$m([i_0 \ldots i_\ell]^j) := p(i_0).P(i_0,i_1).\cdots.P(i_{\ell-1},i_\ell), \tag{2.1}$$

then m extends to a unique probability on X, which we shall also denote by m.

When P is a stochastic matrix over a countable alphabet A, we shall assume that we have been provided with a left-invariant strictly positive probability vector p, and using this we construct the probability (measure) m as before. Note that in general for such matrices P, neither the existence nor the uniqueness of a suitable vector p is guaranteed (see [F]).

From equation (2.1) one sees that the probability m is shift invariant, i.e. $m(S^{-1}(B)) = m(B)$, for all sets $B \in \mathcal{B}$. The process $(X,\mathcal{B},m,S)$ or (simply) S, is called the (2-sided) *Markov shift* defined by the matrix P. We shall also refer to S as a Markov shift over finitely or countably many states (as appropriate), or more precisely, over the (finite or countable) state space A.

2.2. Remarks.

The above definitions are standard, and we refer the reader to [F], [D], or [Se2] for further details. The following definitions are based on those given by A. del Junco and M. Rahe in [J&R], and by W. Parry in [P2].

It is well known (see, for example [F]) that for any $k \times k$ irreducible stochastic matrix P, there exists a unique strictly positive probability vector $p = (p(0),\ldots,p(k-1))$ such that $p.P = p$. Define for each cylinder

$$m([i_0 \ldots i_\ell]^j) := p(i_0).P(i_0,i_1).\cdots.P(i_{\ell-1},i_\ell), \tag{2.1}$$

then m extends to a unique probability on X, which we shall also denote by m.

When P is a stochastic matrix over a countable alphabet A, we shall assume that we have been provided with a left-invariant strictly positive probability vector p, and using this we construct the probability (measure) m as before. Note that in general for such matrices P, neither the existence nor the uniqueness of a suitable vector p is guaranteed (see [F]).

From equation (2.1) one sees that the probability m is shift invariant, i.e. $m(S^{-1}(B)) = m(B)$, for all sets $B \in \mathcal{B}$. The process $(X,\mathcal{B},m,S)$ or (simply) S, is called the (2-sided) *Markov shift* defined by the matrix P. We shall also refer to S as a Markov shift over finitely or countably many states (as appropriate), or more precisely, over the (finite or countable) state space A.

2.2. <u>Remarks.</u>

The above definitions are standard, and we refer the reader to [F], [D], or [Se2] for further details. The following definitions are based on those given by A. del Junco and M. Rahe in [J&R], and by W. Parry in [P2].

2.3. DEFINITIONS.

If (X_1,B_1,m_1,S_1), (X_2,B_2,m_2,S_2) are 2-sided Markov shifts, then a measure preserving isomorphism ϕ between S_1 and S_2 is called *finitary* if for m_1 - a.e. $x = (x_n) \in X_1$, there corresponds a pair of integers (s,t), $s \leq 0 \leq t$, such that for m_1 - a.e. $y \in [x_s x_{s+1} \ldots x_t]^s$, $(\phi(x))_0 = (\phi(y))_0$.

For a cylinder $C = [x_s x_{s+1} \ldots x_t]^s$, $s \leq 0 \leq t$, $x_i \in A$, $s \leq i \leq t$, we define the length of C to be $t-s+1$. We may then define a B_1-measurable function $\ell : X_1 \to \mathbb{N}$ by setting $\ell(x)$ to be the minimum value of $t-s+1$ arising from pairs of integers (s,t) associated to the point $x \in X_1$ in the above definition of finitary. The function ℓ is defined m_1 - a.e., and is called the *code-length* of the finitary isomorphism ϕ.

The finitary isomorphism ϕ is said to have *finite expected code length* when

$$\int_{X_1} \ell(x) dm_1(x) < \infty .$$

The *inverse code length* of ϕ is defined as above with ϕ replaced by ϕ^{-1}. Collectively, the code length and inverse code length of a finitary map ϕ, are referred to as the code lengths of ϕ.

2.4. Remark.

The present definition of finitary is straightforwardly equivalent to that given by W. Parry in [P2].

2.5. DEFINITION.

An invertible measure preserving (resp. non-singular) transformation T from a measure space $(X,\mathcal{B},m)$ to itself is called *reversible* if there exists a measure preserving (resp. non-singular) isomorphism $\phi : X \to X$ between T and T^{-1}, i.e. $\phi T = T^{-1}\phi$ m-a.e..

2.6. Example.

One straightforward example of a reversible transformation is provided by a translation T_g on a compact group G, by a fixed element $g \in G$. A reversing map ϕ for T_g is given by $\phi(h) = h^{-1}, h \in G$. In this example both T_g and ϕ preserve the Haar probability measure on G.

The following result, together with an idea for its proof, was suggested to me by W. Parry.

2.7. THEOREM.

Let S be the Markov shift defined by an irreducible, finite stochastic matrix, then S is reversible by a reversing map ϕ that is measure preserving, finitary, and has finite expected code lengths.

Proof.

The proof will be divided into five parts. We first need to define the reversing map ϕ we shall be considering.

(i) Definition of the reversing map ϕ.

Let X_S denote the doubly infinite product space on which the shift S is defined as the left-shift, i.e.

$$X_S = \prod_{i=-\infty}^{\infty} \{0,1,\ldots,k-1\} \quad ,$$

for some fixed integer $k \geq 2$, and $(Sx)_n = x_{n+1}$, for $n \in \mathbb{Z}$, and $x = (x_n) \in X_S$. (Notice that if $k = 1$, then X_S is a single point, and the theorem is trivial. For this reason, we assume that $k \geq 2$.)

We shall denote the defining matrix for S by P_S , and the σ-algebra and probability measure associated with P_S in Definitions 2.1 by $\mathcal{B}_S$ and m_S respectively. Let us also make similar definitions with respect to the shift S^{-1} , noting that S^{-1} is to be the left-shift on $X_{S^{-1}}$.

Let C_S (resp. $C_{S^{-1}}$) denote the set of cylinders contained in X_S (resp. $X_{S^{-1}}$) of the form

$$[0x_1x_2\ldots x_{r-1}x_r0]^j \quad ,$$

where $x_1,x_2,\ldots,x_r \in \{1,\ldots,k-1\}$, and $j \in \mathbb{Z}$. We define a map $\phi:C_S \to C_{S^{-1}}$ by setting

$$\phi([0x_1x_2\ldots x_{r-1}x_r0]^j) := [0x_rx_{r-1}\ldots x_2x_10]^j \quad . \tag{2.2}$$

The map ϕ can be extended to countable unions and intersections of cylinders from C_S , by defining

$$\phi(\bigcap_{i\in I} C_i) := \bigcap_{i\in I} \phi(C_i) \tag{2.3}$$

and

$$\phi(\bigcup_{i \in I} C_i) := \bigcup_{i \in I} \phi(C_i) \ , \tag{2.4}$$

where I is a countable index set, and $C_i \in C_S$ for all $i \in I$.

Since the set, $\overline{X}_S$, of points $x \in X_S$ whose associated sequence (x_n) contains infinitely many 0's has full measure, i.e.

$$m_S(\{x = (x_n) \in X_S \ ; \ \mathrm{card}\{n \in \mathbb{Z}; (x)_n = 0\} = \infty\}) = 1 \ ,$$

the above map ϕ may be considered as a map from $\overline{X}_S$ to $\overline{X}_{S^{-1}}$, or from X_S to $X_{S^{-1}}$. (Note that for conceptual clarity we carefully distinguish between the spaces on which S and S^{-1} are defined.) Henceforth, we consider ϕ as a map from X_S to $X_{S^{-1}}$; by its definition, ϕ is seen to be B_S-measurable, and to have an inverse ϕ^{-1} which is $B_{S^{-1}}$-measurable, thus ϕ is an isomorphism.

(ii) <u>The map ϕ satisfies $\phi S = S^{-1}\phi$, m_S-a.e.</u>.

By their definitions, the shifts S and S^{-1} satisfy

$$S'(\bigcap_{i \in I} C_i) = \bigcap_{i \in I} S'(C_i) \ , \tag{2.5}$$

$$S'(\bigcup_{i \in I} C_i) = \bigcup_{i \in I} S'(C_i) \ , \tag{2.6}$$

where I is a countable index set such that $C_i \in C_{S'}$ for all $i \in I$, and $S' = S, S^{-1}$. Equations (2.3), (2.4), (2.5), and (2.6) together imply that to prove $\phi S = S^{-1}\phi$ m_S-a.e. on $\overline{X}_S$, it is sufficient to prove $\phi S(C) = S^{-1}\phi(C)$ for all cylinders $C \in C_S$.

Let C be an arbitrary cylinder chosen from $\mathcal{C}_S$, then it is of the form

$$C = [0x_1x_2\ldots x_{r-1}x_r0]^j$$

for some $j \in \mathbb{Z}$, and $x_1, x_2, \ldots, x_r \in \{1,2,\ldots,k-1\}$. We also have

$$\begin{aligned}
\phi S(C) &= \phi([0x_1x_2\ldots x_{r-1}x_r0]^{j-1}) \\
&= [0x_rx_{r-1}\ldots x_2x_10]^{j-1} \text{ ; a cylinder set in } X_{S^{-1}} \\
&= S^{-1}([0x_rx_{r-1}\ldots x_2x_10]^j) \text{ ; since } S^{-1} \text{ is the left-shift on } X_{S^{-1}} \\
&= S^{-1}\phi([0x_1x_2\ldots x_{r-1}x_r0]^j) \\
&= S^{-1}\phi(C) .
\end{aligned}$$

Thus $\phi S = S^{-1}\phi$ m_S-a.e. on $\overline{X}_S$, and since $m_S(\overline{X}_S) = 1$, $\phi S = S^{-1}\phi$ m_S-a.e..

Parts (i) and (ii) together prove that ϕ is an isomorphism between S and S^{-1}.

(iii) <u>The map ϕ is measure preserving.</u>

From the general theory of Markov shifts, the unique strictly positive probability vector p_S associated to P_S in Definitions 2.1 is the same as that for $P_{S^{-1}}$, (see [F; Vol.I, XV.12]) and will be denoted by p. The matrices P_S and $P_{S^{-1}}$ are also related by

$$p(i).P_{S^{-1}}(i,j) = p(j).P_S(j,i) \tag{2.7}$$

for all $i,j \in \{0,1,\dots,k-1\}$. By virtue of equations (2.3) and (2.4), to prove that ϕ is measure preserving, it is sufficient to prove that ϕ preserves the measure of all cylinders $C \in C_S$.

Let $C = [0x_1x_2\dots x_{r-1}x_r0]^j \in C_S$, then by equations (2.1) and (2.2), the following calculation is valid:

$$\begin{aligned}
&m_{S^{-1}}(\phi(C))\\
&= p(0).P_{S^{-1}}(0,x_r).\cdots.P_{S^{-1}}(x_1,0)\\
&= p(0).\left(\frac{p(x_r)}{p(0)}.P_S(x_r,0)\right).\cdots.\left(\frac{p(0)}{p(x_1)}.P_S(0,x_1)\right)\\
&\qquad\qquad ;\ \text{using equation (2.7)}\\
&= p(0).P_S(0,x_1).\cdots.P_S(x_r,0)\\
&= m_S(C)\ .
\end{aligned}$$

The map ϕ is therefore measure preserving.

(iv) <u>The maps ϕ and ϕ^{-1} are finitary</u>.

Every point $x \in \overline{X}_S$ lies in a unique cylinder $C(x) \in C_S$ of the form

$$C(x) = [0i_1i_2\dots i_{r-1}i_r0]^j$$

for some $j \in \mathbb{Z}$, $2\text{-length}(C(x)) \leq j \leq 0$, and $i_1,i_2,\dots,i_r \in \{1,2,\dots,k-1\}$.

Since $m_S(\overline{X}_S) = 1$, the map ϕ is defined m_S-a.e., and hence by equation (2.2), for m_S-a.e. $y \in C(x)$, $(\phi(y))_0 = (\phi(x))_0$. By Definitions 2.3, ϕ is therefore finitary. Proceeding in a similar manner, ϕ^{-1} is also shown to be finitary.

(v) <u>The maps ϕ and ϕ^{-1} have finite expected code length</u>.

To provide estimates for certain recurrence probabilities, the following well-known result will be useful.

<u>Claim</u>.

There exists a constant $0 < \lambda < 1$ such that for all $n \geq \text{card}(A) = k$,

$$m_S(\{x=(x_n)\in X_S;(x)_0=0 \text{ and } (x)_i\neq 0 \text{ for all } i=1,2,\ldots,n\}) \leq p(0).\lambda^n .$$

Proof of claim.

By assumption the defining matrix for S , denoted P_S , is irreducible, and hence for all $i,j \in \{0,1,\ldots,k-1\}$, there exists a smallest positive integer $m = m(i,j)$ such that $P_S^m(i,j) > 0$. However, since this implies that it is possible to get from state i to state j , we can do so in $(k-1)$ steps or less, and therefore $m(i,j) \leq (k-1)$ for all states $i,j \in \{0,1,\ldots,k-1\}$.

Let us define

$$r: = \min_{i,j\in A} \{P_S^{m(i,j)}(i,j)\} , \tag{2.8}$$

and note that $r > 0$. We now define

$$A_0 := \{x = (x_n) \in X_S \; ; \; (x)_0 = 0\}$$

and inductively define the sets A_u , $u = 1,2,\ldots$, by setting

$$A_{u+1} := \{x = (x_n) \in A_u ; (x)_i \neq 0 \text{ for all } i = uk+1,\ldots,(u+1)k\} \; .$$

The Markov property (see, for example [F;Vol.I,XV.13]), together with the definition of r in equation (2.8), implies that

$$m_S(A_{u+1}) \leq (1 - r).m_S(A_u) \; ,$$

for all $u = 0,1,\ldots$, and hence

$$\begin{aligned} m_S(A_u) &\leq (1 - r)^u.m_S(A_0) \\ &= p(0).(1-r)^u \; , \end{aligned} \tag{2.9}$$

for all $u = 1,2,\ldots$.

Choose and fix a constant λ , $0 < \lambda < 1$ such that

$$\lambda^n > (1 - r)^{[n/k]} \; , \quad \text{for all } n = k,k+1,\ldots \quad ,$$

where $[z]$ denotes the integer part of z , then inequality (2.9) implies that for all $n \geq k$,

$$m_S(\{x=(x_n)\in X_S ; (x)_0 = 0 \text{ and } (x)_i \neq 0 \text{ for all } i = 1,\ldots,n\}) \leq p(0).\lambda^n \; .$$

The proof of the claim is complete.

We now proceed with the proof that ϕ and ϕ^{-1} have finite expected code length.

Using the same notation as in part (iv), we define a function $L: X_S \to \mathbb{R}$ by setting

$$L(x) := \text{length}(C(x)) \quad \text{for } x \in \overline{X}_S \ .$$

(In fact we have only defined L on $\overline{X}_S$, but $\overline{X}_S$ has full measure in X_S.) The function L is $\mathcal{B}_S$-measurable and satisfies $L(x) \geq \ell(x)$ m_S-a.e., where ℓ denotes the code length of ϕ, as in Definitions 2.3. To prove that ϕ has finite expected code length, it is therefore sufficient to show that $L \in L^1(X_S, \mathcal{B}_S, m_S)$, this we do as follows.

$$\begin{aligned}
& m_S(\{x \in X_S;\ L(x) = n\}) \\
&= (n-1).m_S(\{x = (x_n) \in X_S;\ (x)_0 = 0 = (x)_{n-1} \text{ and } \\
&\qquad\qquad (x)_i \neq 0 \text{ for all } i = 1,2,\ldots,n-2\}) \\
&\qquad ;\ \text{because the number of cylinders of the correct} \\
&\qquad\ \ \text{form and length } n \text{, is } (n-1) \\
&\leq (n-1).m_S(\{x = (x_n) \in X_S\ ;\ (x)_0 = 0 \text{ and } (x)_i \neq 0 \\
&\qquad\qquad \text{for all } i = 1,2,\ldots,n-2\}) \\
&< (n-1).p(0).\lambda^{n-2}\ , \quad \text{for all } n \geq k+2\ ,
\end{aligned}$$

by the above Claim.

Therefore,

$$\sum_{n=1}^{\infty} m_S(\{x \in X_S \; ; \; L(x) \geq n\})$$

$$= \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} m_S(\{x \in X_S ; \; L(x) = i\})$$

$$\leq \tfrac{1}{2}(k+1)(k+2) + \sum_{n=k+2}^{\infty} \sum_{i=n}^{\infty} p(0).(i-1)\lambda^{i-2}$$

$$; \text{ since for } n = 1,2,\ldots,k+1 \, , \quad m_S(\{x \in X_S ; \; L(x) \geq n\}) \leq 1$$

$$\leq \tfrac{1}{2}(k+1)(k+2) + \sum_{j=k}^{\infty} (j-k+1)(j+1)\lambda^{j}$$

$$< \infty \quad ; \text{ because } 0 < \lambda < 1 \, .$$

Since L is a strictly positive, integer valued function, and $\ell(x) \leq L(x)$ m_S-a.e.,

$$\int_{X_S} \ell(x) dm_S(x) \leq \int_{X_S} L(x) dm_S(x)$$

$$= \sum_{n=1}^{\infty} m_S(\{x \in X_S \; ; \; L(x) \geq n\})$$

$$< \infty \, ,$$

and hence by Definitions 2.3, ϕ has finite expected code length. A similar calculation shows that ϕ^{-1} also has finite expected code length.

The proof of Theorem 2.7 is complete. $\square$

2.8. COROLLARY.

Let S be the Markov shift defined by a finite (not necessarily irreducible) stochastic matrix, and a left-invariant, strictly positive, probability vector; then S is reversible by a reversing map ϕ that is measure preserving, finitary and has finite expected code lengths.

Proof.

Decompose the Markov shift into its (disjoint) irreducible components, A_i (say). By hypothesis, positive measure is given to each irreducible component, and we can therefore find a measure preserving reversing map ϕ_i for each $S|_{A_i}$.

Combining the maps ϕ_i , which are defined on disjoint parts of X_S , produces a reversing map ϕ for S , that preserves the invariant measure arising from the left-invariant vector. By a similar argument to that used in the proof of Theorem 2.7, each map ϕ_i will be finitary and have finite expected code lengths with respect to $S|_{A_i}$ and the appropriate restricted (renormalised) measure. The map ϕ , will inherit these same properties.

□

2.9. Remarks.

Let us define (cf.[P3]) two measure preserving automorphisms T_1, T_2 with preferred partitions α_1, α_2 to be *boundedly equivalent* when there is a measure preserving isomorphism ϕ , $\phi T_1 = T_2 \phi$, such that α_1 and $\phi^{-1}\alpha_2$ boundedly code each other. In [J&R], A. del Junco and M. Rahe proved that finitary codes with finite expected code lengths give rise to bounded codes. This, together with Theorem 2.7 and Corollary 2.8, implies that a (finite state) Markov shift S is boundedly equivalent to its inverse S^{-1}, where here the preferred partition for both S and S^{-1} is the *state partition* that consists of the sets $[i]^0$, for $i \in \{0,1,\ldots,k-1\}$.

Let us denote by $R(n)$ the probability that starting from the state 0, the first re-entry to the state 0, under the shift S, occurs at the n^{th}. step. Similar calculations to those used in the proof of Theorem 2.7 prove that for a (probability) measure preserving Markov shift S defined by a countable irreducible stochastic matrix, the reversing map ϕ defined as above is finitary and has finite expected code lengths if S possesses a state 0 (say) for which

$$\sum_{n=1}^{\infty} n^2 R(n) < \infty .$$

As an example of an irreducible stochastic matrix where the state 0 (and in fact any other state) has finite mean recurrence time, i.e.,

$$\sum_{n=1}^{\infty} nR(n) < \infty , \tag{2.10}$$

but where

$$\sum_{n=1}^{\infty} n^2 R(n) = \infty , \tag{2.11}$$

we provide the following.

2.10. Example.

Define the matrix P_S by

$$P_S := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \cdots & & \\ \frac{1}{8} & 0 & \frac{7}{8} & 0 & \cdots & \\ \frac{1}{27} & 0 & 0 & \frac{26}{27} & 0 & \cdots \\ \frac{1}{4^3} & 0 & 0 & 0 & \frac{4^3-1}{4^3} & \ddots \\ \vdots & \vdots & \vdots & \vdots & 0 & \ddots \\ & & & & \vdots & \end{pmatrix}$$

By inspection of the matrix P_S, for all $n \geq 3$, we have

$$R(n) = \frac{1}{2} \times \frac{7}{8} \times \cdots \times \frac{(n-1)^3-1}{(n-1)^3} \times \frac{1}{n^3}$$

and hence

$$nR(n) = (n-1)R(n-1) \times \left(\frac{(n-1)^3-1}{(n-1)n^2}\right) ,$$

$$n^2R(n) = (n-1)^2R(n-1) \times \left(\frac{(n-1)^3-1}{(n-1)^2n}\right) ,$$

for all $n \geq 3$.

As

$$\frac{nR(n)}{(n-1)R(n-1)} = \frac{(n-1)^3-1}{(n-1)n^2} < \frac{(n-1)^2}{n^2}$$

for $n \geq 3$, comparison with a suitable multiple of the sequence $a_n = \frac{1}{n^2}$ implies inequality (2.10).

To prove inequality (2.11) we first note that

$$\frac{n^2R(n)}{(n-1)^2R(n-1)} = \frac{(n-1)^3-1}{(n-1)^2n} = \frac{(n-1)^3-1}{(n-1)^3} \times \frac{(n-1)}{n}$$

for all $n \geq 3$, and that

$$\prod_{n=3}^{\infty} \left(\frac{(n-1)^3-1}{(n-1)^3}\right) > 0 \quad \text{because} \quad \sum_{n=3}^{\infty} \frac{1}{(n-1)^3} < \infty \ .$$

Since for a fixed $n \geq 3$, and any $m \in \mathbb{N}$,

$$\frac{(n+m)^2R(n+m)}{n^2R(n)} \geq \left(\frac{n}{n+m}\right) \cdot \prod_{i=3}^{\infty} \left(\frac{(i-1)^3-1}{(i-1)^3}\right) ,$$

comparison of the sequence $n^2R(n)$ with a suitable positive multiple of the sequence $a_n = \frac{1}{n}$ implies inequality (2.11).

2.11. Remark.

For some (finite state) Markov shifts a reversing map can be constructed with the stronger property of being a "block isomorphism". Further details of this can be found in Section 5.

THE SET OF REVERSIBLE TRANSFORMATIONS IS RESIDUAL.

Having shown that for finite state Markov shifts a reversing map can always be constructed, we shall now give some indications as to the extent to which the property of reversibility is generic within the classes of non-singular automorphisms, and measure preserving automorphisms on a Lebesgue probability space.

2.12. NOTATION.

Let $(X,\mathcal{B},m)$ be a non-atomic Lebesgue probability space, then by $V = V(X,\mathcal{B},m)$ we shall denote the set of classes of identical (modulo sets of measure zero) non-singular automorphisms of the space $(X,\mathcal{B},m)$. By $U = U(X,\mathcal{B},m)$ we denote those elements of V that preserve the measure m. For reasons of both clarity and brevity, we refer to the elements of U and V as automorphisms. Both of the sets U and V become groups under the operation of composition of automorphisms.

We now define a topology on the sets U and V.

2.13. DEFINITION.

For automorphisms $S,T \in V(X,\mathcal{B},m)$, we define

$$\rho(S,T) := m(\{x \in X;\ S(x) \neq T(x)\}) ,$$

then ρ is a metric, and the topology it defines is called the *uniform topology*. Upon the subset U of V, the uniform topology is defined as the relative topology. We shall use ρ to denote the metric both on U and V.

2.14. Remarks.

The uniform topology is discussed for the measure preserving case by P. Halmos in [H1] and V. Rohlin in [R2]. For the more general situation of non-singular automorphisms, we refer the reader to the book [Fr] by N. Friedman, and the paper [C&F] of R. Chacon and N. Friedman. If we denote by U_ρ (resp. V_ρ) the space U (resp. V) equipped with metric ρ, then (see [R2]) U_ρ (and by a similar argument V_ρ) is a complete metric space. With respect to the uniform topology, both U and V are topological groups.

The automorphisms of both U and V may be further classified as follows.

2.15. DEFINITIONS.

Let Id denote the identity transformation on X. We define an automorphism $T \in V$ to have *period p* if both

$$\rho(T^n, Id) > 0 , \quad \text{for all } 1 \leq n < p$$

and

$$\rho(T^p, Id) = 0 .$$

If T also satisfies

$$\rho(T^n, Id) = 1 , \quad \text{for all } 1 \leq n < p$$

we say it has *strict period p* .

An automorphism T is called *periodic* (resp. *strictly periodic*) if T has period (resp. strict period) p, for some $p \in \mathbb{N}$. The automorphism T is called *aperiodic* if

$$\rho(T^n, Id) = 1 , \quad \text{for all } n = 1,2,\ldots .$$

If T is an automorphism for which there exists a constant $\delta > 0$ such that

$$\rho(T^n, Id) \geq \delta \quad , \quad \text{for all } n = 1,2,\ldots \quad ,$$

we say that T has an *aperiodic component*. If δ is the largest such constant for which the above inequality holds, then T has an *aperiodic component of measure* δ .

2.16. Remarks.

Given an automorphism $T \in V$, the space X can be decomposed into disjoint, measurable, T-invariant sets $X_k = X_k(T)$, $k = 0,1,\ldots$, such that for $k \geq 1$, T has strict period k on X_k , and on X_0 , T is aperiodic.

This decomposition is very useful when approximating transformations in the uniform topology, for more details the reader is referred to [Fr].

The following result indicates to what extent periodic automorphisms may be used to approximate, with respect to the metric ρ , any automorphism from V.

2.17. THEOREM. (Ref. [Fr] and [R2])

Let P denote the set of periodic automorphisms in V , then P (resp. $P \cap U$) is dense in V (resp. U) , with respect to the uniform topology.

2.18. Remarks.

The above result is proved in the measure preserving case, $P \cap U$, by V. Rohlin in [R2]. The non-singular case is due to C. Linderholm, and it is stated in [I]. A proof of this result can be found in [Fr;Th.7.7].

As is to be expected, periodic automorphisms are "well behaved", and in particular we have:

2.19. PROPOSITION.

Non-singular automorphisms with no aperiodic component are reversible.

Proof.

Let T be a non-singular automorphism possessing no aperiodic component, then by Remarks 2.16, we may decompose the space X into disjoint, measurable sets $X_k = X_k(T)$, for $k = 1,2,\dots$, such that on each X_k, T has strict period k. Each set X_k with positive measure can be further decomposed (see [I] and [Fr;Th.7.5]) into k disjoint, measurable sets $X_k^0, X_k^1, \dots, X_k^{k-1}$ satisfying

$$T(X_k^i) = X_k^{i+1 (\mathrm{mod}\ k)}, \quad \text{for all } i = 0,1,\dots,k-1 .$$

The map $\phi : X \to X$ defined by

$$\phi(x) := T^{-2i}(x) \quad \text{when } x \in X_k^i,\ i = 0,1,\dots,k-1 ; \quad k = 1,2,\dots ,$$

is a reversing map for T, that is measure preserving when T is, and thus T is reversible. □

Note that since periodic automorphisms have no aperiodic component, the above result applies to them. Let us now recall the following definitions.

2.20. DEFINITIONS.

In a topological space X, a set $B \subset X$ is called *nowhere dense* if the closure of B in X has empty interior. A countable union of nowhere dense sets is called a *set of first category*, and a set not of first category is said to be of *second category*. A topological space X is called a *Baire space* if and only if for any set $B \subset X$ of first category, the complement $X \backslash B$ is everywhere dense in X. In a Baire space, the complement of any subset of first category is called *residual*. Examples of Baire spaces include complete metric spaces, and locally compact Hausdorff spaces. Note that in a Baire space, any dense G_δ set is a residual set.

With the above definitions in mind, one natural question is: of what category is the set of reversible automorphisms? In answer to this we have:

2.21. THEOREM.

The set of reversible automorphisms in V (resp. U) is residual in V_ρ (resp. U_ρ).

Proof.

Let us denote the set of reversible automorphisms in V by R, and for $\varepsilon > 0$, let A_ε denote the set of automorphisms in V that have an aperiodic component of measure less than ε.

We shall first show that for every $\varepsilon > 0$, the set A_ε is open. For a fixed but arbitrary $\varepsilon > 0$, let $T \in A_\varepsilon$, then if $X_k = X_k(T)$, $k = 0,1,2,...$ is the decomposition of X associated to T in Remarks 2.16, we note that

$$a := m(X_0(T)) < \varepsilon .$$

Choose $K = K(T)$ large enough that

$$m(\bigcup_{k=K}^{\infty} X_k(T)) < (\frac{\varepsilon - a}{3}) ,$$

and then pick $\delta > 0$ small enough that for all $S \in V$, with $\rho(S,T) < \delta$,

$$m(X_k(S)) \geq m(X_k(T)) - (\frac{\varepsilon - a}{3K}) ,$$

for all $k = 1,2,...,K-1$, where $X_k(S)$, $k = 0,1,2,...$ denotes the decomposition of X associated to S. (Such a choice of δ is made possible by the control we have over the measures of the sets $\{x \in X_k(T) ; T^i(x) = S^i(x)$ for all $i = 1,2,...,k\}$, for $k = 1,2,...,K-1$.)

For these choices of K and δ we have

$$m(X_0(S)) \leq 1 - \sum_{k=1}^{K-1} m(X_k(S))$$

$$\leq 1 - (\sum_{k=1}^{K-1} (m(X_k(T)) - (\frac{\varepsilon - a}{3K})))$$

$$\leq 1 + \left(\frac{\varepsilon - a}{3}\right) - \left(1 - m(X_0(T)) - \left(\frac{\varepsilon - a}{3}\right)\right)$$

$$< \varepsilon ,$$

and thus A_ε is open, since we have shown that the open ball of centre T and radius δ (dependent only on T) is contained in A_ε .

We now define

$$P := \bigcap_{n=1}^{\infty}(A_{1/n}),$$

and note that P is a G_δ set. The set P consists, by construction, of all non-singular automorphisms in V with no aperiodic component. By Proposition 2.19, every automorphism in P is reversible, i.e. $P \subset R$. Since periodic automorphisms have no aperiodic component, we see that $P \cap V \subset P$. However, the set $P \cap V$ is dense in V by Theorem 2.17, and thus the set P is a dense G_δ . As V equipped with the metric ρ is a complete metric space, P is a residual set, and hence so too is the set of reversible automorphisms in V .

If we replace V by U everywhere in the above proof, the same argument proves that the set of reversible measure preserving automorphisms is residual in U .

□

2.22. Remark.

Given the above result, another natural question is to ask about the existence of automorphisms that are not reversible. In [A], H. Anzai produced an example of a measure preserving automorphism on the 2-torus that is not conjugate to its inverse, and hence is irreversible. Using this example, we provide a complementary result to Theorem 2.21, as follows.

2.23. THEOREM.

Let $\mathcal{I}$ denote the set of irreversible automorphisms in $\mathcal{V}$, then $\mathcal{I}$ (resp. $\mathcal{I} \cap \mathcal{U}$) is dense in $\mathcal{V}_\rho$ (resp. $\mathcal{U}_\rho$).

Proof.

Let T be a non-singular automorphism in $\mathcal{V}$, and let $\varepsilon > 0$ be chosen arbitrarily. By Theorem 2.17, there exists a periodic automorphism T_1 in $\mathcal{V}$ such that $\rho(T,T_1) < \varepsilon/2$. Since T_1 is periodic, we can find a T_1-invariant set $A \subset X$ with measure strictly between 1 and $1 - \varepsilon/2$.

Denote by S, H. Anzai's (see [A]) irreversible ergodic measure preserving automorphism on the 2-torus, $\mathbb{T}^2$, with Lebesgue probability measure λ. By choosing an isomorphism $\theta : \mathbb{T}^2 \to X\backslash A$ that satisfies

$$m(\theta(C)) = m(X\backslash A).\lambda(C)$$

for all measurable sets $C \subset \mathbb{T}^2$, we can define an automorphism T_2 in $\mathcal{V}$ that contains a "small copy" of H. Anzai's measure preserving irreversible automorphism by setting:

$$T_2(x) := \begin{cases} T_1(x) & \text{for } x \in A \\ \theta S \theta^{-1}(x) & \text{for } x \in X\backslash A . \end{cases}$$

Since ρ is a metric on V, we have

$$\rho(T,T_2) \leq \rho(T,T_1) + \rho(T_1,T_2)$$

$$< \varepsilon/2 + m(X\backslash A) < \varepsilon \quad .$$

The constructed automorphism T_2, is now claimed to be irreversible. To see this, just note that any reversing map ϕ for T_2 must preserve the decomposition of X associated to T_2 in Remarks 2.16. Thus $\theta^{-1}\phi\theta$ would be a reversing map for S on $\mathbb{T}^2$, a contradiction.

In the measure preserving situation, the result is proved by replacing V by U wherever it appears in the above argument. □

2.24. Remark.

Another interesting topology defined on both U and V, is the weak topology (see [H1]). The weak topology on V is the one for which a sequence T_n from V converges to $T \in V$ if and only if for each measurable set E, $m(T_n(E) \Delta T(E)) \to 0$ as $n \to \infty$, where Δ denotes symmetric difference. By using the relative topology, we define the weak topology on U.

Since the uniform topology is finer than the weak topology (see, for example [H1]) all of the density results discussed above are true with respect to the weak topology. However, the Baire category of the set of reversible automorphisms within U or V with the weak topology (although investigated) is not as yet known.

SECTION 3.

AN INFORMATION COCYCLE FOR GROUPS OF NON-SINGULAR TRANSFORMATIONS.

This section is fundamental to the present thesis. Its contents are contained in a joint paper with K. Schmidt ([B&S]), the exposition of which we shall follow very closely. The nature of this joint work is detailed in the Declaration.

The information function is an important and often very useful tool in isomorphism and coding problems of ergodic theory. Its main limitation lies in the fact that, in order to define an information function for an ergodic endomorphism T of a Lebesgue probability space $(X,\mathcal{S},m)$, one has to find a suitable 'invariant' σ-algebra $A \subset \mathcal{S}$ for T, which, by abuse of notation, means that $T^{-1}A \subset A$. Explicit constructions of such σ-algebras are intimately linked with the non-trivial problem of finding generators for T. A much more fundamental difficulty arising from the use of invariant σ-algebras is the imposition of a total order on the collection of σ-algebras $\{T^nA; n \geq 0\}$, or, if T is an automorphism, on $\{T^nA; n \in \mathbb{Z}\}$. This limits the use of information functions to actions of totally ordered semigroups.

In this section we define, for every non-singular endomorphism T of the Lebesgue space $(X,\mathcal{S},m)$, and for every σ-algebra $A \subset \mathcal{S}$ which is not altered too much by T (the precise condition is that the conditional information functions $I_m(A|T^{-1}A)$ and $I_m(T^{-1}A|A)$ are both finite m-a.e.), a function $J_m(A,T)$ with the following properties: $J_m(A,\cdot)$ is a 1-cocycle on the semigroup $\Delta(A)$ of endomorphisms of

$(X,\mathcal{S},m)$ on which it is defined, and $J_m(A,T)$ coincides with the classical information function of T with respect to A if T is measure preserving and $T^{-1}A \subset A$. For arbitrary measure preserving $T \in \Delta(A)$, $J_m(A,T)$ is defined as $I_m(A|T^{-1}A) - I_m(T^{-1}A|A)$. This extension of the classical information function turns out to have a number of interesting properties, and we shall describe some applications in this and later sections.

The section is organised as follows. We first introduce some background material to establish notational and terminological conventions. The following subsection describes an important cohomology on a particular equivalence relation of σ-algebras. This cohomology lies at the root of the information cocycle's cohomological properties. In the next subsection we define the information cocycle and prove some of its basic properties. Within the following subsection we discuss a connection between $J_m(A,T)$ and the entropy $h(T)$ of a measure preserving transformation T . In the final subsection we use the information cocycle, together with some coding results from Section 1, to give a simple proof of a well known formula for the entropy of a C^2 Anosov diffeomorphism.

3.1. <u>NOTATION AND BACKGROUND.</u>

For convenience of presentation, throughout this section we shall use (X,S,m) to denote a Lebesgue probability space.

Unless we have reason to be particularly careful we shall identify functions which coincide m-a.e.. If A and B are sub-σ-algebras of S we say that $A = B \pmod m$ if $\sup_{A\in A} \inf_{B\in B} m(A\Delta B) = \sup_{B\in B} \inf_{A\in A} m(A\Delta B) = 0$. In particular we choose and fix a countably generated σ-algebra $S_0 \subset S$ with $S_0 = S \pmod m$ and such that (X,S_0) is a standard Borel space. For a given sub-σ-algebra A of S, $\hat{A}$ will stand for the measurable partition of X given by A, and $\{m_x^A; x\in X\}$ the associated decomposition of m into conditional measures. Since we shall have to manipulate these objects very carefully we shall describe them in a little more detail.

We choose a countably generated σ-algebra $A_0 \subset S_0$ with $A_0 = A \pmod m$ and define an equivalence relation on X by calling two points $x_1, x_2 \in X$ equivalent, if and only if, for every $A \in A_0$, $x_1 \in A \iff x_2 \in A$. The equivalence class of a point $x \in X$ with respect to this equivalence relation is denoted by $[x]_A$, and $\hat{A}$ is defined as the measurable partition $\{[x]_A; x\in X\}$ of X. The sets $[x]_A$, $x \in X$, are referred to as the *atoms* of A. If p is a probability measure on (X,S) equivalent to m $(p \sim m)$, the partition $\hat{A}$ induces a decomposition of p into conditional measures: there exists a family $\{p_x^A; x\in X\}$ of probability measures on (X,S_0) with the following properties.

(i) For every $x \in X$, $p_x^A([x]_A) = 1$. (3.1)

(ii) For every $B \in S_0$, the map $x \to p_x^A(B)$ is A_0-measurable, and

$$p(B) = \int p_x^A(B)dp(x) \ . \tag{3.2}$$

The objects $\hat{A}$ and $\{p_x^A; x \in X\}$ are essentially unique and independent of the choice of S_0 and A_0. Furthermore, if $f \in L^1(X,S,p)$, and if $E_p(f|A)$ denotes the conditional expectation of f, given A, with respect to the measure p, then

$$E_p(f|A)(x) = \int f \, dp_x^A \quad \text{p-a.e.} \ . \tag{3.3}$$

The uniqueness properties of conditional measures imply that, for m-a.e. $x \in X$, the measures p_x^A and m_x^A are equivalent, and

$$\frac{dp_x^A}{dm_x^A} = \frac{dp}{dm} \Big/ E_m\left(\frac{dp}{dm}\Big|A\right) \quad m_x^A\text{-a.e.} \ . \tag{3.4}$$

If we denote by p_A and m_A the restrictions of p and m (respectively) to the σ-algebra A, then $E_m(\frac{dp}{dm}|A)$ is equal to the Radon-Nikodym derivative $\frac{dp_A}{dm_A}$ on (X,A). In particular we get, for equivalent probability measures m,p,q on (X,S),

$$E_p\left(\frac{dq}{dp}\Big|A\right) . E_m\left(\frac{dp}{dm}\Big|A\right) = E_m\left(\frac{dq}{dm}\Big|A\right) \ . \tag{3.5}$$

If $B \subset S$ is another sub-σ-algebra, we define $A \vee B$ to be the refinement $\{A \cap B; A \in A, B \in B\}$.

3.2. DEFINITION.

Let A, B be sub-σ-algebras of S, and let p be a probability measure equivalent to m on (X,S), then the *conditional information* of B, given A, with respect to the measure p, is defined by

$$I_p(B|A)(x) := -\log p_x^A([x]_{A\vee B}),\ x \in X. \tag{3.6}$$

3.3. Remark.

An alternative, but equivalent, definition of conditional information is contained in [P3].

Clearly, $I_p(B|A) < \infty$ p-a.e. if and only if, for p-a.e. $x \in X$, $p_x^A([x]_{A\vee B}) > 0$ or, equivalently, if $\hat{B}$ partitions p-a.e. element of $\hat{A}$ (essentially) into a countable number of subsets. Another way of expressing this is to say that p-a.e. measure $p_x^{A\vee B}$ is absolutely continuous with respect to p_x^A. In this case we have, for p-a.e. $x \in X$,

$$\log \frac{dp_x^{A\vee B}}{dp_x^A}(y) = I_p(B|A)(y), \tag{3.7}$$

for p_x^A-a.e. $y \in [x]_{A\vee B}$, as a consequence of the uniqueness of conditional measures.

For notational convenience and brevity, we shall make the following definition.

3.4. DEFINITION.

For p a probability measure on (X,S) satisfying $p \ll m$, and

A a sub-σ-algebra of S , we define

$$I(p,m|A) := -\log E_m(\frac{dp}{dm}|A) \ .$$

The next result relates the conditional information functions arising from two equivalent measures.

3.5. LEMMA.

Let (X,S,m) be a Lebesgue probability space, and p be a probability measure equivalent to m . For sub-σ-algebras $A,B \subset S$,

$$I_p(B|A) = I_m(B|A) + I(p,m|A \vee B) - I(p,m|A) \ . \qquad (3.8)$$

Proof.

Combine equations (3.4) and (3.7), and use Definition 3.4.

□

We now investigate how these expressions are transformed by a non-singular endomorphism T of (X,S,m) .

The straightforward proof of the following lemma is omitted.

3.6. LEMMA.

Let T be a non-singular endomorphism of (X,S,m) , and let A be a sub-σ-algebra of S . For every measurable function $f:X \to \mathbb{R}$ for which $f \circ T$ is m-integrable, we have

$$E_m(f\circ T|T^{-1}A) = E_{mT^{-1}}(f|A)\circ T \quad . \tag{3.9}$$

3.7. Remark.

If A is a sub-σ-algebra of S, and T is a non-singular endomorphism, the atoms of $T^{-1}A$ can be chosen to satisfy

$$[x]_{T^{-1}A} = T^{-1}[Tx]_A \quad , \quad x \in X \; , \tag{3.10}$$

and this together with equation (3.3) and Lemma 3.6, yields

$$(mT^{-1})^A_{Tx} = (m_x^{T^{-1}A})T^{-1} \; , \tag{3.11}$$

for m-a.e. $x \in X$.

3.8. LEMMA.

Let A,B be sub-σ-algebras of S, and p be a probability measure on (X,S) equivalent to m, then

a) $I(p,m|T^{-1}A) = I(pT^{-1},mT^{-1}|A)\circ T$

b) $I_m(T^{-1}B|T^{-1}A) = I_{mT^{-1}}(B|A)\circ T$.

Proof.

a) Noting Definition 3.4, we apply Lemma 3.6 to the function $f = \dfrac{dpT^{-1}}{dmT^{-1}}$, and obtain

$$E_m(\frac{dp}{dm}|T^{-1}A) = E_m(\frac{dpT^{-1}}{dmT^{-1}}\circ T|T^{-1}A) = E_{mT^{-1}}(\frac{dpT^{-1}}{dmT^{-1}}|A)\circ T \ ,$$

as required.

b) The following calculation proves the result.

$I_m(T^{-1}B|T^{-1}A)(x)$

$= -\log m_x^{T^{-1}A}([x]_{T^{-1}A\vee T^{-1}B})$; by Definition 3.2

$= -\log m_x^{T^{-1}A}(T^{-1}[Tx]_{A\vee B})$; from equation (3.10)

$= -\log (mT^{-1})_{Tx}^{A}([Tx]_{A\vee B})$; using equation (3.11)

$= I_{mT^{-1}}(B|A)(Tx)$

for m-a.e. $x \in X$.

□

A COHOMOLOGY.

We shall now describe a cohomology on a certain equivalence relation of σ-algebras. This cohomology is fundamental to the properties of the information cocycle that we shall later investigate.

3.9. NOTATION.

Let (X,S,m) be a Lebesgue probability space, and let $\Omega(S)$ denote the collection of all sub-σ-algebras of S .

3.10. DEFINITION.

Two σ-algebras $A, B \in \Omega(S)$ are called I-*related* (denoted by $A \underset{I}{\sim} B$) if

$$I_m(A|B) + I_m(B|A) < \infty \quad m\text{-a.e.}\ .$$

3.11. PROPOSITION.

The relation $\underset{I}{\sim}$ does not depend on the choice of m within its equivalence class. Furthermore, $\underset{I}{\sim}$ is an equivalence relation on $\Omega(S)$.

Proof.

The first assertion follows from Lemma 3.5. Symmetry and reflexivity of $\underset{I}{\sim}$ are immediate, and its transitivity is an easy consequence of Remark 3.3.

□

A general recipe for constructing cohomologies from equivalence relations is contained in [F&M], and this construction can, of course, be applied to the pair $(\Omega(S), \underset{I}{\sim})$. Rather than go into this in any generality, we shall instead consider one particular example.

3.12. NOTATION.

Let $U(X,m,\mathbb{R})$ denote the additive group (under pointwise addition) of m-equivalence classes of measurable real valued functions on X . For every integer $n \geq 1$, define $R^{(n)} \subset \Omega(S)^{n+1}$ by

$$R^{(n)} = \{(A_0,\dots,A_n) \in \Omega(S)^{n+1} \ ; \ A_{k-1} \underset{I}{\sim} A_k \ \text{ for } k = 1,\dots,n\} \quad ,$$

and put $R^{(o)} = \Omega(S)$. Let $C^{-1} = U(X,m,\mathbb{R})$ and denote by C^n , $n \geq 0$, the group of all maps from $R^{(n)}$ into $U(X,m,\mathbb{R})$ with pointwise addition as composition. The group C^n is called the n-th cochain group. For $n \geq 0$ we have coboundary operators $\delta_{n-1} : C^{n-1} \to C^n$ given by

$$(\delta_{n-1}f)(A_0,\dots,A_n) = \sum_{k=0}^{n} (-1)^k f(A_0,\dots,\overset{\Delta}{A_k},\dots,A_n) \quad ,$$

where Δ indicates the term to be omitted. As usual one puts $Z^n := \ker \delta_n$, $B^n := \delta_{n-1}(C^{n-1})$ and $H^n := Z^n/B^n$ for $n \geq 0$, and calls these groups the group of *n-cocycles, n-coboundaries* and the *n-th cohomology group,* respectively. Our reason for introducing these cohomology groups is that they provide the framework for the following observation which is crucial to the rest of this section.

Define, for every probability measure $p \sim m$, a function $K_p : R^{(1)} \to U(X,m,\mathbb{R})$ by

$$K_p(A,B) := I_p(A|B) - I_p(B|A) \tag{3.12}$$

for every $A,B \in \Omega(S)$ with $A \underset{I}{\sim} B$.

3.13. <u>THEOREM</u>.

For every $A,B,C \in \Omega(S)$ with $A \underset{I}{\sim} B \underset{I}{\sim} C$, and for every probability measure p on (X,S) with $p \sim m$ we have

$$K_p(A,B) + K_p(B,C) = K_p(A,C) . \tag{3.13}$$

In other words, $K_p \in Z^1$.

Proof.

$$K_p(A,B) + K_p(B,C) - K_p(A,C)$$

$$= I_p(A|B) + I_p(C|A\vee B) - I_p(C|B) - I_p(A|B\vee C)$$

$$+ I_p(B|C) + I_p(A|B\vee C) - I_p(A|C) - I_p(B|A\vee C)$$

$$+ I_p(C|A) + I_p(B|A\vee C) - I_p(B|A) - I_p(C|A\vee B)$$

$$= \sum_{D\in\{A,B,C\}} I_p(A\vee B\vee C|D) - I_p(A\vee B\vee C|D)$$

$$= 0 .$$

Note that this computation is valid since, for

$$A,B,C \in \Omega(S) , A \underset{I}{\gamma} B \underset{I}{\gamma} C \Rightarrow A \vee B \underset{I}{\gamma} C .$$

□

We now show that K_p does not depend essentially on the choice of p within the equivalence class of m .

3.14. PROPOSITION.

For every $A,B \in \Omega(S)$ with $A \underset{I}{\sim} B$,

$$K_p(A,B) = K_m(A,B) + I(p,m|A) - I(p,m|B) \ . \tag{3.14}$$

In other words, K_p and K_m are cohomologous, i.e. $K_p - K_m \in B^1$.

Proof.

Equation (3.14) follows from Lemma 3.5. To see that the map sending the pair $(A,B) \in R^{(1)}$ to $I(p,m|A) - I(p,m|B)$ is a coboundary, just note that it is the image under δ_0 of the map that sends $A \in R^{(0)}$ to $I(p,m|A)$. □

3.15. Remark.

In passing we derive two more expressions which will be useful later. Using equation (3.7), K_p can be written in terms of conditional measures as

$$K_p(A,B)(x) = \log \frac{dp_x^A}{dp_x^B}(x) \tag{3.15}$$

for m-a.e. $x \in X$. Equation (3.15) gives another "quick" proof of the cocycle equation (3.13).

For the next statement we assume T to be a non-singular endomorphism of (X,S,m) . Using Lemma 3.8 (part b) together with equations

(3.12) and (3.14), we obtain

$$K_m(T^{-1}A,T^{-1}B)$$

$$= K_{mT^{-1}}(A,B)\circ T$$

$$= K_m(A,B)\circ T + I(mT^{-1},m|A)\circ T - I(mT^{-1},m|B)\circ T \ . \qquad (3.16)$$

A different interpretation of equation (3.13) leads to a 1-cocycle for certain group actions on (X,S,m) in the sense of [F&M] and [Sc] which extends directly the concept of the classical information function in ergodic theory, and which has a number of interesting properties.

THE INFORMATION COCYCLE.

3.16. NOTATION.

Let (X,S,m) be a non-atomic Lebesgue probability space and let $U(X,m,\mathbb{R})$ be defined as in Notation 3.12. We denote by $\mathrm{End}(X,S,m)$ the semigroup of all non-singular endomorphisms of (X,S,m), and write $\mathrm{Aut}(X,S,m)$ for the subgroup of all invertible elements in $\mathrm{End}(X,S,m)$.

For the rest of this subsection we choose and fix a sub-σ-algebra A of S. Using this σ-algebra we shall define a distinguished sub-semigroup $\Delta(A) \subset \mathrm{End}(X,S,m)$ and an information cocycle $J_m(A,\cdot): \Delta(A) \to U(X,m,\mathbb{R})$.

The first step in defining the information cocycle consists of describing its domain. For the definition of the equivalence relation $\underset{I}{\sim}$ we refer to Definition 3.10.

3.17. DEFINITION.

$$\Delta(A) := \{T \in \mathrm{End}(X,S,m);\ T^{-1}A \underset{I}{\sim} A\}. \tag{3.17}$$

$$\Gamma(A) := \Delta(A) \cap \mathrm{Aut}(X,S,m)\ . \tag{3.18}$$

Clearly $\Delta(A)$ is a semigroup and $\Gamma(A)$ a group. Both depend only on the measure class of m, and we have $\Delta(A) = \Delta(B)$, and hence $\Gamma(A) = \Gamma(B)$, whenever $B \subset S$ is a σ-algebra satisfying $B \underset{I}{\sim} A$.

3.18. DEFINITION.

Let p be a probability measure on (X,S) equivalent to m, and let $T \in \Delta(A)$. Recalling the definition of $K_p(A,T^{-1}A)$ in equation (3.12) and of $I(pT^{-1},p|A)$ in Definition 3.4, we put

$$J_p(A,T) := K_p(A,T^{-1}A) + I(pT^{-1},p|A)\circ T\ . \tag{3.19}$$

The next two results will justify the name 'information cocycle' for the function $J_m(A,\cdot)$.

3.19. <u>THEOREM</u>.

For every $S,T \in \Delta(A)$,

$$J_m(A,S)\circ T + J_m(A,T) = J_m(A,ST) \ . \qquad (3.20)$$

Proof.

$$J_m(A,S)\circ T + J_m(A,T)$$

$$= K_m(A,S^{-1}A)\circ T + I(mS^{-1},m|A)\circ ST$$

$$+ K_m(A,T^{-1}A) + I(mT^{-1},m|A)\circ T \quad ; \text{ by Definition 3.18}$$

$$= K_m(T^{-1}A,T^{-1}S^{-1}A) + I(mT^{-1},m|S^{-1}A)\circ T$$

$$+ I(mS^{-1},m|A)\circ ST + K_m(A,T^{-1}A) \quad ; \text{ from equation (3.16)}$$

$$= K_m(A,T^{-1}A) + K_m(T^{-1}A,T^{-1}S^{-1}A)$$

$$+ I(mT^{-1}S^{-1},mS^{-1}|A)\circ ST + I(mS^{-1},m|A)\circ ST \ ; \text{ using Lemma 3.8 (part a)}$$

$$= K_m(A,T^{-1}S^{-1}A) + I(mT^{-1}S^{-1},m|A)\circ ST \quad ; \text{ by Theorem 3.13, and equation (3.5) with Definition 3.4}$$

$$= J_m(A,ST) \ .$$

□

3.20. <u>THEOREM</u>.

Let p be a probability measure on (X,S) equivalent to m , $B \subset S$ be a sub-σ-algebra with $B \underset{I}{\sim} A$, and let $T \in \Delta(A)$, then

$$J_p(B,T) = J_m(A,T) + \phi \circ T - \phi \ , \qquad (3.21)$$

where

$$\phi = K_m(A,B) - I(p,m|B) \ . \qquad (3.22)$$

Proof.

The following calculations prove the result.

$J_p(B,T)$

$= K_p(B,T^{-1}B) + I(pT^{-1},p|B)\circ T$; by Definition 3.18

$= K_m(B,T^{-1}B) + I(p,m|B) - I(p,m|T^{-1}B) + I(pT^{-1},p|B)\circ T$

; from Proposition 3.14

$= J_m(B,T) - I(mT^{-1},m|B)\circ T$

$+ I(p,m|B) - I(p,m|T^{-1}B) + I(pT^{-1},p|B)\circ T$

$= J_m(B,T) - I(mT^{-1},m|B)\circ T$

$- I(pT^{-1},mT^{-1}|B)\circ T - I(p,pT^{-1}|B)\circ T$

$+ I(p,m|B)$

; using Lemma 3.8 (part a), and equation (3.5) with Definition 3.4

$= J_m(B,T) + I(p,m|B) - I(p,m|B)\circ T \qquad (3.23)$

; from equation (3.5) and Definition 3.4.

We also have

$$J_m(B,T) = K_m(B,T^{-1}B) + I(mT^{-1},m|B)\circ T$$

$$= K_m(B,A) + K_m(A,T^{-1}A) + K_m(T^{-1}A,T^{-1}B) + I(mT^{-1},m|B)\circ T$$

; by Theorem 3.13

$$= K_m(B,A) + K_m(A,T^{-1}A) + K_m(A,B)\circ T$$

$$+ I(mT^{-1},m|A)\circ T - I(mT^{-1},m|B)\circ T + I(mT^{-1},m|B)\circ T$$

; from equation (3.16)

$$= J_m(A,T) + K_m(A,B)\circ T - K_m(A,B) .$$

Combining this with equation (3.23), we obtain equations (3.21) and (3.22), as required.

□

The following definition uses the terminology of [F&M] and [Sc].

3.21. DEFINITION.

A map $a:\Gamma(A) \to U(X,m,\mathbb{R})$ is a 1-*cocycle* if

$$a(S)\circ T + a(T) = a(ST) \tag{3.24}$$

for every $S,T \in \Gamma(A)$. A 1-cocycle a is called a *coboundary* if there exists a map $b \in U(X,m,\mathbb{R})$ with

$$a(S) = b\circ S - b , \tag{3.25}$$

for all $S \in \Gamma(A)$, and two cocycles a_1 and a_2 are said to be *cohomologous* if $a_1 - a_2$ is a coboundary.

3.22. Remarks.

Theorem 3.19 shows that the restriction of $J_m(A,\cdot)$ to $\Gamma(A)$ is a 1-cocycle, and Theorem 3.20 states that $J_m(A,\cdot)$ and $J_p(B,\cdot)$ are cohomologous whenever $m \sim p$ and $A \underset{I}{\sim} B$. The fact that the relations (3.20) and (3.21) extend to $\Delta(A)$ is not just an empty generalization, since traditionally both the information function, and its integral, entropy, have been associated with measure preserving ergodic endomorphisms as much as with automorphisms. We shall discuss the connection between $J_m(A,\cdot)$ and the classical information function in more detail in the next subsection. First, however, we shall derive an alternative expression for the information cocycle.

3.23. Remark.

If $T \in \Gamma(A)$, we can re-write $J_m(A,T)$ as follows:

$$
\begin{aligned}
&J_m(A,T)(x) \\
&= K_m(A,T^{-1}A)(x) + I(mT^{-1},m|A)(Tx) \\
&= \log \frac{dm_x^A}{dm_x^{T^{-1}A}}(x) - \log E_m\left(\frac{dmT^{-1}}{dm}\middle|A\right)(Tx)
\end{aligned}
$$

; by equation (3.15) and Definition 3.4

$$= \log \frac{dm_x^A}{d((mT^{-1})_{Tx}^A T)}(x) \cdot \frac{dm_A}{d(mT^{-1})_A}(Tx) \quad m\text{-a.e.} \tag{3.26}$$

; using equation (3.11) and the remark preceding equation (3.5).

In the special case where $T \in \Gamma(A)$ preserves m we get

$$J_m(A,T)(x) = \log \frac{dm_x^A}{d(m_{Tx}^A)T}(x) \quad m\text{-a.e.} \; . \tag{3.27}$$

3.24. Remark.

If $S,T \in \Gamma(A)$ preserve the measure m, then equation (3.20) is an immediate consequence of the cocycle equation (3.13) (or, in fact, of equation (3.27) and the chain rule for Radon-Nikodym derivatives).

We conclude this subsection with a technical point. Let $G \subset \Gamma(A)$ be a countable group and let $[G] = \{V \in \mathrm{Aut}(X,S,m); Vx \in Gx$ for m-a.e. $x \in X\}$ denote the full group of G (see [Dy], [Kri1]).

3.25. THEOREM.

1) $[G] \subset \Gamma(A)$.

2) For every $V \in [G]$ and $g \in G$ we have

$$J_m(A,V)(x) = J_m(A,g)(x) \tag{3.28}$$

m-a.e. on the set $B_g = \{x \in X; Vx = gx\}$.

Proof.

Let $V \in [G]$, and $g \in G$. If $m(B_g) > 0$ we have, for m-a.e. $x \in B_g$

$$m_x^A([x]_{A \vee V^{-1}A}) \geq m_x^A(g^{-1}([gx]_A \cap gB_g)) > 0$$

from equation (3.10), and this shows that $I_m(V^{-1}A|A) < \infty$ m-a.e. (cf. equation (3.6)). Using equation (3.11) one can similarly prove that $I_m(A|V^{-1}A) < \infty$ m-a.e.. Having proved part 1) we note that equation (3.28) is equivalent to

$$J_m(A,g^{-1}V) = 0 \quad m\text{-a.e. on } B_g \tag{3.29}$$

for every $g \in G$. The proof of this theorem is thus completed by the following lemma.

3.26. LEMMA.

Let $T \in \Gamma(A)$ and put $N_T := \{x \in X; Tx = x\}$, then $J_m(A,T) = 0$ m-a.e. on N_T .

Proof.

Assume that $m(N_T) > 0$ - otherwise the assertion is trivial. Equations (3.26) and (3.4) imply that, for m-a.e. $x \in N_T$,

$$J_m(A,T)(x)$$

$$= \log \frac{dm_x^A}{d(mT^{-1})_x^A}(x) \cdot \frac{dm_A}{d(mT^{-1})_A}(x)$$

$$= \log \frac{dm}{dmT^{-1}}(x) = 0 \ .$$

□

3.27. Remark.

Theorem 3.25 has several useful implications. If $G \subset \Gamma(A)$ is a countable group, it expresses $J_m(A,V)$ in terms of $J_m(A,g)$, $g \in G$, for every $V \in [G]$. As a consequence of equation (3.28) we see that, for $V \in [G]$ and $x \in X$, $J_m(A,V)(x)$ depends only on the pair of points (x,Vx) and not on what V does to the rest of the space (we are ignoring m-null sets at the moment). For a much more detailed discussion of this type of cocycle we refer to [F&M] and [Sc]. The first assertion in Theorem 3.25 also shows that $\Gamma(A)$ is a very large group. In the most interesting case, where m_A is non-atomic and where m-a.e. m_x^A, $x \in X$, is non-atomic, Theorem 3.25, together with H.A. Dye's theorem [Dy] implies that $\Gamma(A)$ contains isomorphic copies of, for example, all measure preserving ergodic automorphisms of (X,S,m).

3.28. Remark.

If $T \in \Gamma(A)$ satisfies $T^{-1}A \subset A$, the information cocycle reduces to

$$J_m(A,T) = I_m(A|T^{-1}A) - \log E_m\left(\frac{dmT^{-1}}{dm}\middle|A\right) \circ T \ ,$$

and this expression already appears in [P5;Lemma 1.2] (with a minor misprint). In Section 4 we shall study this situation in more detail.

Theorem 3.20 is a generalization of [P2;Corollary 1].

In the following subsection we shall discuss the connection between the information cocycle $J_m(A,T)$ and the entropy $h(T)$ of T .

ENTROPY AND THE INFORMATION COCYCLE.

3.29. NOTATION.

Let (X,S,m) be a non-atomic Lebesgue probability space and denote by $End^1(X,S,m)$ and $Aut^1(X,S,m)$ the set of measure preserving elements in $End(X,S,m)$ and $Aut(X,S,m)$, respectively.

In this subsection we reverse the point of view adopted so far within this section and consider the collection of all sub-σ-algebras which allow us to define information cocycles for a fixed semigroup $G \subset End(X,S,m)$.

3.30. DEFINITION.

If $G \subset End(X,S,m)$ is a semigroup we put $\Sigma(G) = \{A \subset S;\ A$ is a σ-algebra with $g^{-1}A \underset{I}{\gamma} A$ for every $g \in G\}$. For $G \subset End^1(X,S,m)$ we define

$$\Sigma^1(G) := \{A \in \Sigma(G);\ \int (I_m(A|g^{-1}A) + I_m(g^{-1}A|A))dm < \infty \text{, for every } g \in G\} .$$

If $T \in \mathrm{End}(X,S,m)$ we shall write $\Sigma(T)$ instead of $\Sigma(\{T,T^2,T^3,\ldots\})$. Similarly, if $T \in \mathrm{End}^1(X,S,m)$, $\Sigma^1(T)$ will stand for $\Sigma^1(\{T,T^2,T^3,\ldots\})$.

3.31. Remark.

For a fixed endomorphism $T \in \mathrm{End}^1(X,S,m)$, one of the commonly used definitions of the entropy $h(T)$ of T is the following (cf. [P1;Theorem 5.14]):

$$h(T) := \sup_{\substack{A \in \Sigma^1(T) \\ A \subset T^{-1}A}} \int I_m(A|T^{-1}A)\,dm . \tag{3.30}$$

Bearing in mind that, for $A \in \Sigma^1(T)$ with $A \subset T^{-1}A$,

$$J_m(A,T) = I_m(A|T^{-1}A) ,$$

equation (3.30) can be rewritten as

$$h(T) := \sup_{\substack{A \in \Sigma^1(T) \\ A \subset T^{-1}A}} \int J_m(A,T)\,dm . \tag{3.31}$$

Our next proposition allows us to express $h(T)$ more naturally in terms of the information cocycle.

3.32. PROPOSITION.

For $T \in \mathrm{End}^1(X,S,m)$ we have

$$h(T) = \sup_{A \in \Sigma^1(T)} \int J_m(A,T)dm . \tag{3.32}$$

Proof.

It suffices to consider the case where $h(T) < \infty$. From equation (3.30) we know that $h(T) \le \sup_{A\in\Sigma^1(T)} \int J_m(A,T)dm$. In order to establish the converse inequality we fix $A \in \Sigma^1(T)$. By Theorem 3.20 we have, for every integer $n \ge 1$,

$$J_m(A \vee \ldots \vee T^{-n}A,T) = J_m(A,T) + \phi_n \circ T - \phi_n ,$$

where ϕ_n is easily seen to lie in $L^1(X,S,m)$. Hence

$$\int J_m(A,T)dm = \int J_m(A\vee\ldots\vee T^{-n}A,T)dm$$
$$\le \int I_m(A|T^{-1}A\vee\ldots\vee T^{-n-1}A)dm , \tag{3.33}$$

for every $n \ge 1$. Since $I_m(A|T^{-1}A) \in L^1(X,S,m)$, the last term in inequality (3.33) decreases to $\int I_m(A|T^{-1}A^-)dm$, where $A^- := \bigvee_{k=0}^{\infty} T^{-k}A$ (cf.[P1;Theorem 5.5]) , and equation (3.30) yields

$$\int J_m(A,T)dm \le \int I_m(A|T^{-1}A^-)dm$$
$$= \int I_m(A^-|T^{-1}A^-)dm \le h(T) .$$

□

For the remainder of this subsection we assume T to be a measure preserving automorphism of (X,S,m) . If T has finite entropy we can give a precise expression for $\int J_m(A,T)dm$ for every $A \in \Sigma^1(T)$.

3.33. THEOREM.

Let T be a measure preserving automorphism of (X,S,m) with finite entropy, and let $A \in \Sigma^1(T)$. Put

$$A^+ := \bigvee_{k=0}^{\infty} T^k A \tag{3.34}$$

and

$$A^- := \bigvee_{k=0}^{\infty} T^{-k} A , \tag{3.35}$$

then

$$\int J_m(A,T)dm = \int I_m(A^-|T^{-1}A^-)dm - \int I_m(A^+|TA^+)dm . \tag{3.36}$$

Proof.

Using Theorem 3.20 and the fact that $A \underset{I}{\sim} A \vee T^{-1}A \vee \dots \vee T^{-n}A$ for every integer n we get

$$\int J_m(A,T)dm = \int J_m(A \vee T^{-1}A \vee \ldots \vee T^{-n+1}A, T)dm$$

$$= \int I_m(A|T^{-1}A \vee \ldots \vee T^{-n}A)dm - \int I_m(T^{-n}A|A \vee \ldots \vee T^{-n+1}A)dm$$

$$= \int I_m(A|T^{-1}A \vee \ldots \vee T^{-n}A)dm - \int I_m(A|TA \vee \ldots \vee T^{n}A)dm \ .$$

Letting $n \to \infty$ and using a martingale theorem of entropy theory due to Chung (see [P1;Theorem 5.5]) we obtain equation (3.36).

□

3.34. <u>COROLLARY</u>.

$\int J_m(A,T)dm = h(T)$ if and only if

$h(T) = \int I_m(A^-|T^{-1}A^-)dm$ and $A^+ = TA^+$.

3.35. <u>COROLLARY</u>.

Let C be a finite algebra with

$$C_T := \bigvee_{k=-\infty}^{\infty} T^k C = S$$

and let $A \in \Sigma^1(T)$ satisfy

$$C \subset A \subset C^- \tag{3.37}$$

and

$$A^+ = TA^+ , \tag{3.38}$$

then

$$\int J_m(A,T)dm = h(T) .$$

Proof.

From condition (3.37) it is clear that $A^- = C^-$ and that $I_m(A|T^{-1}A^-) \geq I_m(C|T^{-1}C^-)$. Since $C_T = S$, by [P1;Corollary 3.12] we also have

$$\begin{aligned} h(T) &= \int I_m(C|T^{-1}C^-)dm \\ &\leq \int J_m(A,T)dm . \end{aligned}$$

Now apply Proposition 3.32.

□

3.36. Remark.

Even if $A \in \Sigma^1(T)$ satisfies $\int J_m(A,T)dm = h(T)$, it need not be the case that $A \underset{J}{\sim} A^-$.

3.37. COROLLARY.

If $T \in \mathrm{Aut}^1(X,S,m)$ has zero entropy then

$$\int J_m(A,T)dm = 0 \quad \text{for every } A \in \Sigma^1(T) .$$

We conclude this subsection with an instructive example.

3.38. Example.

Let $Y := \prod_{k=-\infty}^{\infty} \{0,1\}$, $p := \prod_{k=-\infty}^{\infty} p_k$, with $p_k(0) = p_k(1) = 1/2$ for all $k \in \mathbb{Z}$, $\mathcal{T}$ the product σ-algebra, and S the Bernoulli shift on $(Y,\mathcal{T},p)$ given by $(S(y))_n = y_{n+1}$ for all $n \in \mathbb{Z}$ and every $y = (y_n) \in Y$. We define C to be algebra generated by the state partition, i.e. $C := \{\emptyset, Y, \{y; y_0 = 0\}, \{y; y_0 = 1\}\}$, and put $C^- := \bigvee_{k=0}^{\infty} S^{-k}C$, $C^+ := \bigvee_{k=0}^{\infty} S^k C$. Next put $(X,\mathcal{S},m) := (Y\times Y, \mathcal{T}\times\mathcal{T}, p\times p)$, $T := S\times S$ and $A := C^-\times C^+$. The transformation T is the Bernoulli shift on four symbols with equal probabilities, and thus $h(T) = \log 4 \neq 0$. The σ-algebra A satisfies the following conditions:

a) $A \in \Sigma^1(T)$.

b) $A_T = \mathcal{S}$, and $\bigcap_{k=-\infty}^{\infty} T^k A = \mathcal{N} = \{\emptyset, X\}$,

where $A_T := \bigvee_{k=-\infty}^{\infty} T^k A$.

c) $J_m(A,T) = 0$.

THE INFORMATION COCYCLE AND THE ENTROPY OF ANOSOV MAPS.

In this subsection we shall use the information cocycle to prove a well known formula for the entropy of Anosov diffeomorphisms. The proof we shall give utilises some of the coding results from Section 1, but a similar argument avoiding the use of these results is also possible (see [B&S]). For further details of both notation and definitions we refer to Section 1 and [Bo4].

The result for which we shall given an alternative proof appears in [Bo4], and we rephrase its statement as follows.

3.39. THEOREM.

Let T be a C^2 Anosov diffeomorphism of a compact, connected, C^∞ Riemannian manifold X, preserving a smooth probability measure m, then

$$h(T) = \int_X \log|\det(DT|_{E_x^u})|dm(x) ,$$

where $DT|_{E_x^u}$ is the restriction of the derivative DT of T at x to the tangent space at x of the unstable manifold of T through x.

Proof.

First note that Anosov diffeomorphisms satisfy Axiom A^*, and we shall therefore be able to apply results from Section 1.

Let $\delta = \delta(T) > 0$ be the constant appearing in Notation 1.9, that corresponds to T. Choose and fix a smooth finite partition α with $\text{diam}(\alpha) < \delta$, then with respect to T, α generates the Borel σ-algebra S, i.e. $\alpha_T = \bigvee_{i=-\infty}^{\infty} T^{-i}\alpha = S$, and thus (see [P1; Corollary 3.12]),

$$\int J_m(\alpha^-,T)dm = h(T) .$$

Let β denote the measurable partition whose elements consist of the connected components (of stable manifolds) that result from intersecting the partition α with the collection of all stable manifolds in X .

Since $\operatorname{diam}(\alpha) < \delta$, the elements of the partition α^- are pieces of stable manifold, and hence $\beta \subset \alpha^-$. This immediately implies that $I_m(\beta|\alpha^-) = 0$ m-a.e..

To prove that $I_m(\alpha^-|\beta) < \infty$ m-a.e., we apply Corollary 1.14 together with a result of W. Parry [P3; Theorem 3] (which although only stated for a finite partition β , is however valid for general measurable partitions). This argument actually shows more, it shows that $I_m(\alpha^-|\beta) \in L^p(X,\mathcal{S},m)$, for all $1 \le p < \infty$.

For later convenience we shall find it useful to denote $\hat{\beta}$ by $\mathcal{B}$, and with this notation we have now proved that $\hat{\alpha}^- \underset{I}{\sim} \mathcal{B}$.

In [Pu&Sh] it is shown that C^2 Anosov diffeomorphisms have smooth measures on the leaves of the stable (and unstable) foliation, and hence for m-a.e. $x \in X$, the measure $m_x^{\mathcal{B}}$ is equivalent to the leaf measure λ_x induced by the Riemannian structure of X on the element of the partition β containing x . The measures λ_x , $x \in X$, also satisfy

$$\log \frac{d\lambda_x}{d\lambda_{Tx}T}(x) = -\log\left|\det\left(DT\Big|_{E_x^s}\right)\right| , \tag{3.39}$$

where E_x^s denotes the tangent space of the stable manifold of T at x .

Using the above preliminary results, we can make the following calculation, which will complete the proof.

$$\int \log|\det(DT|_{E_x^u})|\,dm(x)$$

$$= -\int \log|\det(DT|_{E_x^s})|\,dm(x) \qquad ;\ \text{since } T \text{ preserves } m$$

$$= \int \log \frac{d\lambda_x}{d\lambda_{Tx}T}(x)\,dm(x) \qquad ;\ \text{using equation (3.39)}$$

$$= \int \log \left(\frac{d\lambda_x}{dm_x^\beta}(x)\,\frac{dm_x^\beta}{dm_{Tx}^\beta T}(x)\,\frac{dm_{Tx}^\beta T}{d\lambda_{Tx}T}(x)\right)dm(x)$$

; the chain-rule for Radon-Nikodym derivatives

$$= \int \log \frac{dm_x^\beta}{dm_{Tx}^\beta T}(x)\,dm(x)$$

; because both $-\log|\det(DT|_{E_x^s})|$ and $\log \dfrac{dm_x^\beta}{dm_{Tx}^\beta T}(x)$ are integrable functions

$$= \int J_m(\beta,T)\,dm \qquad ;\ \text{from equation (3.27)}$$

$$= \int J_m(\alpha^-,T) \qquad ;\ \text{as } \hat{\alpha}^- \underset{I}{\sim} \beta$$

$$= h(T)\ .$$

□

3.40. <u>Remark</u>.

When T is only a C^1 Anosov diffeomorphism of X preserving m, the above proof still shows that

$$h(T) = \int J_m(\mathcal{B},T)dm \ .$$

The following proposition shows that for hyperbolic toral automorphisms the cohomology class of the information cocycle associated to the "past σ-algebra" of a small, smooth, finite partition is very closely related to entropy.

3.41. PROPOSITION.

Let T be a hyperbolic toral automorphism of the n-torus X, with Borel σ-algebra $\mathcal{S}$ and normalised Haar measure m. There exists a constant $\delta = \delta(T) > 0$ such that for any finite Lipschitz partition α, with $\mathrm{diam}(\alpha) < \delta$,

$$J_m(\alpha^-,T) = h(T) + g - g \circ T \ ,$$

for some function $g \in L^p(X,\mathcal{S},m)$, for all $1 \le p < \infty$.

Proof.

Since hyperbolic toral automorphisms satisfy Axiom A^*, and preserve Haar measure, we shall be able to apply results from Section 1.

Let $\delta = \delta(T) > 0$ be the constant associated to T in Notation 1.9, and note that as a consequence of Corollary 4.12, it is sufficient to prove the result for just one Lipschitz partition α, with $\mathrm{diam}(\alpha) < \delta$.

By considering the n-torus as the unit n-cube with opposite faces identified, it is straightforward to construct two measurable partitions α,β with the following properties:

1) α is a finite, smooth partition with $\mathrm{diam}(\alpha) < \delta$.

2) Each element of β consists of a single connected component of a stable manifold, and all such elements are translationally congruent.

3) $\mathrm{diam}(\beta) < \delta$.

4) The partitions α and β satisfy $\beta \subset \alpha^{-}$.

For such partitions α,β we can now apply a similar argument to that used in the proof of Theorem 3.39, to show that $I_{\underline{m}}(\beta|\alpha^{-}) = 0$ m-a.e., and $I_{\underline{m}}(\alpha^{-}|\beta) \in L^{p}(X,\mathcal{S},m)$, for all $1 \le p < \infty$. To complete the proof we use Theorem 3.20, noting that condition 2 implies both $\hat{\beta} \underset{I}{\curlyvee} T^{-1}\hat{\beta}$, and $J_{\underline{m}}(\beta,T) = h(T)$.

□

SECTION 4.

THE INFORMATION COCYCLE OF SINGLE TRANSFORMATIONS.

Having investigated the information cocycle for groups of non-singular transformations in Section 3, in this section we shall restrict our attention to the information cocycle associated to a single transformation (and its powers).

We first consider to what extent the measure determines the L^p-spaces in which the information cocycle lies. Then, for an ergodic automorphism T, and T-invariant σ-algebras A, C we study conditions that imply $T^{-1}A \underset{I}{\overset{\sim}{\vee}} A \underset{I}{\overset{\sim}{\vee}} C \underset{I}{\overset{\sim}{\vee}} T^{-1}C$. As a consequence of some of this work together with coding results from Section 1, we are then able to canonically associate an information cocycle to measure preserving Axiom A^* homeomorphisms of compact metric spaces.

Finally, motivated by the work on R. Fellgett and W. Parry [F&P] on "regular pairs" of σ-algebras, we look at another condition between two σ-algebras A and C that also imples $T^{-1}A \underset{I}{\overset{\sim}{\vee}} A \underset{I}{\overset{\sim}{\vee}} C \underset{I}{\overset{\sim}{\vee}} T^{-1}C$.

Throughout this section $(X,\mathcal{B},m)$ will denote a Lebesgue probability space, and we shall make frequent use of notations and definitions from the previous sections.

Our main tools are Theorem 3.20, together with the following lemmas and their corollaries. As indicated, several of these results are due to W. Parry.

4.1. LEMMA. (Ref. [P2;Lemma 3])

For sub-σ-algebras $\mathcal{A},\mathcal{C} \subset \mathcal{B}$, $d_m(\mathcal{A},\mathcal{C}) < 2$ if and only if $I_m(\mathcal{A}|\mathcal{C}) < \infty$ on a set of positive measure.

4.2. COROLLARY.

For p a probability measure on $(X,\mathcal{B})$ equivalent to m, and sub-σ-algebras $\mathcal{A},\mathcal{C} \subset \mathcal{B}$, $d_p(\mathcal{A},\mathcal{C}) < 2$ if and only if $d_m(\mathcal{A},\mathcal{C}) < 2$.

Proof.

Combine Lemma 3.5 with Lemma 4.1.

□

4.3. LEMMA.

Let $\mathcal{A},\mathcal{C}$ and $\mathcal{D}$ be sub-σ-algebras of $\mathcal{B}$, such that $I_m(\mathcal{A}|\mathcal{C}) < \infty$ m-a.e. and $I_m(\mathcal{C}|\mathcal{D}) < \infty$ m-a.e., then $I_m(\mathcal{A}|\mathcal{D}) < \infty$ m-a.e..

Proof.

The result follows straightforwardly from Remark 3.3.

□

4.4. LEMMA. (Cf. [P2;Corollary to Proposition 1])

If T is a non-singular endomorphism of the Lebesgue probability space $(X,\mathcal{B},m)$, and α is a finite partition, then the function $I_m(\alpha|T^{-1}\alpha^-) \in L^p(X,\mathcal{B},m)$, for all $1 \le p < \infty$.

4.5. COROLLARY.

If α is a finite partition of the Lebesgue probability space $(X,\mathcal{B},m)$, and T is a non-singular endomorphism of X for which there exists a constant $c > 0$ satisfying

$$\frac{1}{c} \le \frac{dmT^{-1}}{dm} \le c \qquad m\text{-a.e.},$$

then the information cocycle $J_m(\alpha^-,T) \in L^p(X,\mathcal{B},m)$, for all $1 \le p < \infty$.

As can be seen from the following proposition together with Corollary 4.5, to some extent it is the sub-σ-algebra A that determines the L^p-spaces to which the information cocycle belongs, provided we are prepared to consider other measures within the measure class of m. More precisely, we have:

4.6. PROPOSITION.

Let T be a non-singular automorphism of the Lebesgue probability space $(X,\mathcal{B},m)$, then for any $\epsilon > 0$, there exists a probability measure $p \sim m$ satisfying

$$1 - \varepsilon < \frac{dpT^{-1}}{dp} < 1 + \varepsilon \qquad m\text{-a.e..}$$

Proof.

Choose a real number c such that

$$1 - \varepsilon < \frac{1}{c} < 1 < c < 1 + \varepsilon \ ,$$

and for all sets $B \in \mathcal{B}$ define

$$p(B) := \left(\frac{c-1}{c+1}\right) \cdot \sum_{i=-\infty}^{\infty} c^{-|i|} m(T^{i}B) \ . \qquad (4.1)$$

By the Vitali-Hahn-Saks Theorem (see [Fr; Theorem 3.8]), p is a measure. It is also a probability measure equivalent to m, for which

$$1 - \varepsilon < \frac{dpT^{-1}}{dp} < 1 + \varepsilon \qquad m\text{-a.e..}$$

□

4.7. Remark.

If m is a T-invariant measure, then the measure p constructed in equation (4.1) is equal to m.

We now describe more precisely to what extent the cohomology class of the information cocycle $J_m(A,T)$ depends on A, for T-invariant sub-σ-algebras A; that is for sub-σ-algebras A that satisfy $T^{-1}A \subset A$. Recall from Remark 3.28, that for such σ-algebras the information cocycle has the simplified form:

$$J_m(A,T) = I_m(A|T^{-1}A) - \log E_m(\frac{dmT^{-1}}{dm}|A)\circ T ,$$

or using Definition 3.4,

$$J_m(A,T) = I_m(A|T^{-1}A) + I(mT^{-1},m|A)\circ T .$$

4.8. <u>THEOREM</u>. (Cf. [P2;Corollary 1])

Let T be an ergodic non-singular automorphism of the non-atomic Lebesgue probability space $(X,\mathcal{B},m)$, and let A,C be sub-σ-algebras such that $T^{-1}A \subset A$, $T^{-1}C \subset C$, and $D_m(A,C) < 2$. If $I_m(A|T^{-1}A) < \infty$ m-a.e., then $J_m(A,T)$ and $J_m(C,T)$ differ by a finite measurable function of the form $\phi\circ T - \phi$, and all of the functions $I_m(C|T^{-1}C)$, $I_m(A|C)$, $I_m(C|A)$ are finite m-a.e., hence $T^{-1}A \underset{\mathrm{I}}{\gamma} A \underset{\mathrm{I}}{\gamma} C \underset{\mathrm{I}}{\gamma} T^{-1}C$.

Proof.

Provided all of the terms appearing are finite m-a.e., Theorem 3.20 implies that

$$J_m(A,T) = J_m(A\vee C,T) + I_m(C|A)\circ T - I_m(C|A) . \qquad (4.2)$$

To prove that these functions are finite m-a.e., first note that from the condition $D_m(A,C) < 2$ and Lemma 4.1, we can conclude that the function $I_m(C|A)$ is finite on a set of positive measure. Let us define the sets

$$E: = \{x \in X \; ; \; I_m(C|A)(x) < \infty \},$$

$$F: = \{x \in X \; ; \; J_m(A \vee C,T)(x) < \infty \}$$

$$= \{x \in X \; ; \; I_m(A \vee C|T^{-1}(A \vee C))(x) < \infty \},$$

where the above equality of sets follows from the finiteness m-a.e. of the function $I(mT^{-1},m|A \vee C) \circ T$. By assumption, $I_m(A|T^{-1}A) < \infty$ m-a.e., and hence $J_m(A,T) < \infty$ m-a.e.. As a consequence, from equation (4.2), we have

$$E = F \cap T^{-1}E \subset T^{-1}E .$$

The ergodicity of T , together with the finiteness of $I_m(C|A)$ on a set of positive measure, therefore implies that $m(E) = 1 = m(F)$, and thus both $I_m(C|A)$ and $J_m(A \vee C,T)$ are finite m-a.e..

Reversing the roles of A and C , Theorem 3.20 similarly implies that

$$J_m(C,T) = J_m(A \vee C,T) + I_m(A|C) \circ T - I_m(A|C) , \tag{4.3}$$

whenever all of the terms appearing are finite m-a.e.. Again using Lemma 4.1 and the condition $D_m(A,C) < 2$, we can also prove that the function $I_m(A|C)$ is finite on a set of positive measure.

Now define the sets

$$E' := \{x \in X \; ; \; I_m(A|C)(x) < \infty\} \,,$$

$$F' := \{x \in X \; ; \; J_m(C,T)(x) < \infty\} \,.$$

Since $J_m(A \vee C,T) < \infty$ m-a.e., equation (4.3) implies that

$$T^{-1}E' = E' \cap F' \subset E' \,.$$

The ergodicity of T^{-1}, together with the finiteness of $I_m(A|C)$ on a set of positive measure, shows that $m(E') = 1 = m(F')$, and thus the finiteness m-a.e. of the functions $I_m(A|C)$, $J_m(C,T)$ and $I_m(C|T^{-1}C)$.

Combining equations (4.2) and (4.3), all the terms of which have now been proved finite m-a.e., we obtain:

$$J_m(A,T) = J_m(C,T) + K_m(A,C) - K_m(A,C) \circ T, \tag{4.4}$$

and this completes the proof.

□

4.9. <u>NOTATION</u>.

If T is a non-singular automorphism of the Lebesgue probability space $(X,\mathcal{B},m)$, then by a slight abuse of notation, we shall (as is usual) call a real valued measurable function of the form $\phi \circ T - \phi$, a

T-coboundary (or a *coboundary with respect to* T), and two functions which differ by a T-coboundary will be called *T-cohomologous*.

4.10. COROLLARY.

Let T be an ergodic non-singular automorphism of the non-atomic Lebesgue probability space $(X,\mathcal{B},m)$, and let A,C be sub-σ-algebras such that $T^{-1}A \subset A$, $T^{-1}C \subset C$, and for which there exist positive integers M and N with $d_m(A,T^M C) < 2$ and $d_m(C,T^N A) < 2$. If $I_m(A|T^{-1}A) < \infty$ m-a.e., then $T^{-1}A \underset{I}{\sim} A \underset{I}{\sim} C \underset{I}{\sim} T^{-1}C$ and $J_m(A,T)$ is T-cohomologous to $J_m(C,T)$.

Proof.

The following calculation is valid provided all of the terms that appear are finite m-a.e..

$$
\begin{aligned}
& I_m(A|T^{-1}A) + I_m(T^{-N}C|A) \\
& = I_m(A \vee T^{-N}C|T^{-1}A) \\
& = I_m(A \vee T^{-N}C \vee T^{-N-1}C|T^{-1}A) \\
& = I_m(A \vee T^{-N}C|T^{-N-1}C \vee T^{-1}A) \\
& \quad + I_{mT^{-1}}(T^{-N}C|A)\circ T \qquad (4.5)
\end{aligned}
$$

; from Lemma 3.8 (part b).

By combining Lemma 3.8 (part b) with Lemma 4.1 and Corollary 4.2, we see that $d_m(C,T^NA) < 2$ precisely when $d_m(T^{-N}C,A) < 2$. The condition $d_m(C,T^NA) < 2$, together with Lemma 4.1, therefore implies that $I_m(T^{-N}C|A) < \infty$ on a set of positive measure, and by using the same method as in the proof of Theorem 4.8, we can then apply the ergodicity of T to prove the finiteness m-a.e. of all of the functions in equation (4.5). (Note that here we have used the fact that $I_{mT^{-1}}(T^{-N}C|A) < \infty$ when and only when $I_m(T^{-N}C|A) < \infty$; this follows from Lemma 3.5.)

Provided all the terms that appear are finite m-a.e., we also have

$$\begin{aligned} &I_m(C|T^{-1}C) + I_m(T^{-M}A|C) \\ &= I_m(C\vee T^{-M}A|T^{-1}C\vee T^{-M-1}A) \\ &\quad + I_{mT^{-1}}(T^{-M}A|C)\circ T . \end{aligned} \tag{4.6}$$

To prove that all of these terms are finite m-a.e., we proceed as follows.

From above, $I_m(A\vee T^{-N}C|T^{-N-1}C\vee T^{-1}A) < \infty$ m-a.e., and using Lemma 3.8 together with Lemma 3.5 we obtain $I_m(T^NA\vee C|T^{-1}C\vee T^{N-1}A) < \infty$ m-a.e.. This implies that

$$I_m(T^{-M}A\vee C|T^{-1}C\vee T^{N-1}A) < \infty \quad \text{m-.a.e..} \tag{4.7}$$

Since $I_m(A|T^{-1}A) < \infty$ m-a.e., by repeatedly using Lemma 4.3, we have

$I_m(T^{N-1}A|T^{-M-1}A) < \infty$ m-a.e., and therefore (as a consequence of Remark 3.3)

$$I_m(T^{-1}C\vee T^{N-1}A|T^{-1}C\vee T^{-M-1}A) < \infty \quad \text{m-a.e..} \tag{4.8}$$

Combining inequalities (4.7) and (4.8) via Lemma 4.3, we get

$$I_m(T^{-M}A\vee C|T^{-1}C\vee T^{-M-1}A) < \infty \quad \text{m-a.e..} \tag{4.9}$$

We now use the same technique as in the proof of Theorem 4.8 (and above) to imply from the condition $d_m(A,T^MC) < 2$, and the ergodicity of T^{-1}, that all of the terms in equation (4.6) are finite m-a.e..

Using Lemma 4.3 repeatedly, the finiteness m-a.e. of $I_m(A|T^{-1}A)$ and $I_m(T^{-M}A|C)$ implies that $I_m(A|C) < \infty$ m-a.e.. Similarly, the finiteness m-a.e. of $I_m(C|T^{-1}C)$ and $I_m(T^{-N}C|A)$ shows that $I_m(C|A) < \infty$ m-a.e.. We have therefore proved that $T^{-1}A \underset{\mathrm{I}}{\gamma} A \underset{\mathrm{I}}{\gamma} C \underset{\mathrm{I}}{\gamma} T^{-1}C$. From this and Lemma 4.1, it is immediate that $D_m(A,C) < 2$.

To prove that $J_m(A,T)$ and $J_m(C,T)$ are T-cohomologous, and thus complete the proof, we apply Theorem 4.8.

□

The next result relates the concept of ϵ-bounded code (see Definition 1.1) to the T-cohomology class of the information cocycle.

4.11. COROLLARY. (Cf. [P3;Corollary to Theorem 1])

Let T be an ergodic non-singular automorphism of the non-atomic Lebesgue probability space $(X,\mathcal{B},m)$. If α,β are finite partitions that ε-boundedly code each other for some $0 \le \varepsilon < 2$, then $J_m(\alpha^-,T)$ and $J_m(\beta^-,T)$ are T-cohomologous.

Proof.

It is easily seen that

a) $T^{-1}\alpha^- \subset \alpha^-$, $T^{-1}\beta^- \subset \beta^-$.

From Lemma 4.4 we also have

b) $I_m(\alpha|T^{-1}\alpha^-) < \infty$ m-a.e., $I_m(\beta|T^{-1}\beta^-) < \infty$ m-a.e..

By assumption, α ε-boundedly codes β, for some $0 \le \varepsilon < 2$, and hence there exists a $k \in \mathbb{N}$ for which

$$d_m(\alpha_k^{n+k},\beta_0^{n+2k}) \le \varepsilon < 2 \text{, for all } n \in \mathbb{N} .$$

This implies that

c) $d_m(T^{-k}\alpha^-,\beta^-) \le \varepsilon < 2$.

Similarly, there exists an $\ell \in \mathbb{N}$ such that

d) $d_m(T^{-\ell}\beta^-,\alpha^-) < 2$.

Using properties a,b,c and d, we can apply Corollary 4.10, and thus complete the proof.

□

We now show that it is possible to associate an information cocycle, in a "canonical" manner, to Axiom A^* homeomorphisms. For definitions of some of the terms used, we refer to Section 1.

4.12. <u>COROLLARY</u>.

Let (X,ρ) be a compact metric space, m be a non-atomic Borel probability measure on X, and $f:X \to X$ be an ergodic homeomorphism that satisfies Axiom A^* and preserves m. There exists a constant $\delta > 0$ such that if α,β are finite Lipschitz partitions with $\mathrm{diam}(\alpha) < \delta$, and $\mathrm{diam}(\beta) < \delta$, then $J_m(\alpha^-,f)$ and $J_m(\beta^-,f)$ differ by a function of the form $\phi\circ f - \phi$, where (denoting the Borel σ-algebra by $\mathcal{B}$) $\phi \in L^p(X,\mathcal{B},m)$, for all $1 \le p < \infty$.

Proof.

Let δ be the constant appearing in the proof of Proposition 1.13, and note that the term "small partitions" in Proposition 1.13, includes all partitions with diameter less than δ. By combining Proposition 1.13 (with its proof), Corollary 1.15, and Corollary 4.11, the information cocycles $J_m(\alpha^-,f)$ and $J_m(\beta^-,f)$ are seen to be f-cohomologous, i.e. to differ by a function of the form $\phi\circ f - \phi$. From equation (4.4) of Theorem 4.8 we also have $\phi = I_m(\beta^-|\alpha^-) - I_m(\alpha^-|\beta^-)$.

To complete the proof we require to show that $\phi \in L^p(X,\mathcal{B},m)$ for all $1 \le p < \infty$, but this follows from Corollary 1.14, used together with a result of W. Parry (see [P3;Theorem 3]).

□

Next, we investigate another situation where it is possible to produce a cocycle-coboundary equation for the information cocycles of two sub-σ-algebras. The conditions relating the two sub-σ-algebras were motivated by the work of R. Fellgett and W. Parry on "regular pairs" of sub-σ-algebras (see [Fe], [F&P]), and we shall give an extension of one of their results as a corollary to the following theorem.

4.13. THEOREM.

Let T be a non-singular automorphism of the non-atomic Lebesgue probability space $(X,\mathcal{B},m)$, and let A,C be sub-σ-algebras of $\mathcal{B}$ for which there exist integers M,N with $I_m(C|T^M A) < \infty$ m-a.e., and $I_m(A|T^N C) < \infty$ m-a.e.. If $A \underset{I}{\gamma} T^{-1}A$, then $A \underset{I}{\gamma} C \underset{I}{\gamma} T^{-1}C$ and $J_m(A,T)$ is T-cohomologous to $J_m(C,T)$.

Proof.

By combining Lemma 3.8 (part b) with equation(3.8), we can conclude from the condition $I_m(A|T^N C) < \infty$ m-a.e., that

$$I_m(T^{-N-1}A|T^{-1}C) < \infty \quad \text{m-a.e..} \tag{4.10}$$

Applying Lemma 4.3 repeatedly, a similar argument shows that

$I_m(A|T^{-1}A) < \infty$ m-a.e. implies

$$I_m(T^M A|T^{-N-1}A) < \infty \text{ m-a.e..} \tag{4.11}$$

Using conditions (4.10) and (4.11), together with $I_m(C|T^M A) < \infty$ m-a.e. and Lemma 4.3, we now prove that $I_m(C|T^{-1}C) < \infty$ m-a.e..

The same technique proves that $I_m(T^{-1}C|C) < \infty$ m-a.e., and hence (by Definition 3.10) $C \underset{I}{\sim} T^{-1}C$.

In a similar way one can also show the finiteness m-a.e. of both functions $I_m(A|C)$ and $I_m(C|A)$. Thus the sub-σ-algebras A and C satisfy $A \underset{I}{\sim} C$.

To complete the proof, we apply Theorem 3.20.

□

4.14. COROLLARY. (Cf. [Fe;Theorem 2 of Section 1])

Let T be a non-singular automorphism of the non-atomic Lebesgue probability space (X,B,m) , and let A,C be sub-σ-algebras such that $T^{-1}A \subset A$, $T^{-1}C \subset C$ and for which there exists an integer M with $C \subset T^M A$, and $A \subset T^M C$. If $I_m(A|T^{-1}A) < \infty$ m-a.e., then $J_m(A,T)$ and $J_m(C,T)$ differ by a function of the form $\phi \circ T - \phi$; furthermore, when $I_m(A|T^{-1}A) \in L^\infty(X,B,m)$, the function $\phi \in L^\infty(X,B,m)$.

Proof.

The conditions $C \subset T^M A$ and $A \subset T^M C$ straightforwardly imply that $I_m(C|T^M A) = I_m(A|T^M C) = 0$ m-a.e., and hence by using $I_m(A|T^{-1}A) < \infty$

m-a.e., Theorem 4.13 proves that $J_m(A,T)$ is T-cohomologous to $J_m(C,T)$, and that $T^{-1}A \underset{I}{\sim} A \underset{I}{\sim} C \underset{I}{\sim} T^{-1}C$. From Theorem 3.20 we therefore have

$$J_m(A,T) = J_m(C,T) - K_m(A,C)\circ T + K_m(A,C) ,$$

i.e. $\phi = -K_m(A,C)$.

To prove the final conclusion, we need to estimate $K_m(A,C)$ in terms of $I_m(A|T^{-1}A)$. This we do as follows:

$$K_m(A,C)$$

$$= K_m(A,T^{-M}A) + K_m(T^{-M}A,C) \text{ ; by Theorem 3.13 with } A \underset{I}{\sim} T^{-M}A \underset{I}{\sim} C$$

$$= I_m(A|T^{-M}A) - I_m(C|T^{-M}A) \text{ ; using } T^{-M}A \subset A \text{ and } T^{-M}A \subset C \ .$$

As $A \subset T^MA$ and $C \subset T^MA$, the fact that $I_m(T^MA|T^{-M}A) \in L^\infty(X,B,m)$ whenever $I_m(A|T^{-1}A) \in L^\infty(X,B,m)$ shows that both of the functions $I_m(A|T^{-M}A)$ and $I_m(C|T^{-M}A)$ are also in $L^\infty(X,B,m)$, and this completes the proof.

□

4.15. <u>Remarks</u>.

When T is measure preserving, the proof just given shows that with the hypotheses of Corollary 4.14, if $I_m(A|T^{-1}A) \in L^p(X,B,m)$, for $1 \le p \le \infty$, then $\phi \in L^p(X,B,m)$.

The conditions relating the sub-σ-algebras A and C in Corollary 4.14 occur naturally within the setting of "regular isomorphisms" between transformations. For results concerning such isomorphisms between measure preserving transformations we refer to [P&T2].

With slight modifications, many of the preceding results can be extended to include the possibility of an infinite σ-finite measure, as a consequence of the following observation.

4.16. Observation.

Let $(X,\mathcal{B},m)$ be an infinite σ-finite Lebesgue measure space, A be a sub-σ-algebra of $\mathcal{B}$, and G be a group of non-singular automorphisms of $(X,\mathcal{B},m)$.

Select a finite, strictly positive, measurable function f, that satisfies

$$\int_X f dm = 1 ,$$

(such functions exist) and define a probability measure p on $(X,\mathcal{B})$ equivalent to m by putting

$$p(B) := \int_B f dm ,$$

for all sets $B \in \mathcal{B}$.

All probability measures constructed in this way are mutually equivalent, thus if

$$G \subset \Gamma_p(A) = \{T \in \mathrm{Aut}(X,\mathcal{B},p); T^{-1}A \underset{I}{\sim} A\},$$

then we can define the (G-) cohomology class of the information cocycle associated to A and $T \in G$ to be the (G-) cohomology class of $J_p(A,T)$.

By considering cohomology classes of functions, instead of functions, it is therefore possible to work within the setting of infinite σ-finite Lebesgue measure spaces.

SECTION 5.

THE FORM OF THE INFORMATION COCYCLE AND SOME INVARIANTS OF THE RELATED COCYCLE-COBOUNDARY EQUATION.

Our aim in the first subsection of this section is to calculate the form of the information cocycle (with respect to the past σ-algebra) of generalised Markov shifts, and finite co-ordinate changes on shift spaces. The expression produced for finite co-ordinate changes has recently been found to provide an invariant of finitary isomorphisms with finite expected coding lengths (see [P&S]).

The second subsection concerns invariants of the information cocycle-coboundary equation. For block isomorphisms between measure preserving Markov shifts over countably many states we produce an invariant that can distinguish between measure preserving Bernoulli shifts over countably many states whose associated probability vectors are not permutations of each other. This invariant is then shown to be closely related to the zeta function to be found in [P&T1]. A straightforward example is given to show that it is possible for "different" Markov shifts to be block isomorphic. Next, we produce two invariants of the cocycle-coboundary equation in which all the functions that appear are essentially bounded. These invariants are "calculated" for measure preserving Markov shifts over finitely many states, and then related to the $\beta(t)$ function described in [T1]. Finally, we produce some examples of Bernoulli shifts that although neither block isomorphic nor "regularly isomorphic", are not distinguishable using any presently known invariants of other types of isomorphism.

5.1. Remark and Notation.

Let $X = \prod_{i=-\infty}^{\infty} \{0,1,\ldots,k-1\}$, and let $T:X \to X$ be the (left-) shift transformation defined by $(Tx)_n := x_{n+1}$, for all $n \in \mathbb{Z}$ and $x = (x_n) \in X$. Upon the σ-algebra $\mathcal{B}$ generated by all cylinder sets, we shall define a probability measure m with respect to which T will be called a (generalised) Markov shift, or (generalised) Bernoulli shift. The probability m is obtained from a system of stochastic matrices and probability vectors as follows.

5.2. DEFINITION.

For each integer n , let P_n be a $k \times k$ stochastic matrix, and let p_n be a strictly positive probability vector of length k . Assume also, that the matrices and row-vectors P_n, p_n, $n \in \mathbb{Z}$, satisfy:

a) $p_n P_n = p_{n+1}$, for all $n \in \mathbb{Z}$.

b) $P_n(i,j) > 0$ if and only if $P_0(i,j) > 0$, for all $n \in \mathbb{Z}$ and all $i,j \in \{0,1,\ldots,k-1\}$.

We now define a (generalised) Markov probability measure m on $(X,\mathcal{B})$, by defining the measure of a cylinder set $C = [x_{-i}x_{-i+1}\ldots x_{j-1}x_j]^{-i}$ to be

$$m(C) := p_{-i}(x_{-i}) \cdot P_{-i}(x_{-i},x_{-i+1}) \cdot \cdots \cdot P_{j-1}(x_{j-1},x_j) \quad .$$

As a consequence of Kolmogorov's consistency theorem (ref. [Pa; Chapter V]),

m can be extended to a probability measure (also denoted by m) on the full σ-algebra $\mathcal{B}$. The measure m will be called a *(generalised) Markov measure* on X, and the shift T will be called a *(generalised) Markov shift* on X whenever it is non-singular with respect to m. If in the above we had also required that all of the matrices P_n were equal and that the vectors p_n were the same, we would have obtained a measure preserving Markov shift, as defined in Definitions 2.1.

5.3. DEFINITION.

For all integers n, let P_n and p_n be matrices and row-vectors (respectively) that satisfy the conditions of Definition 5.2. If the further condition

c) $P_n(i,j) = p_{n+1}(j)$, for all $n \in \mathbb{Z}$ and all $i,j \in \{0,1,\dots,k-1\}$

is also satisfied, then the measure m constructed as in Definition 5.2 will be called a *(generalised) Bernoulli measure* on X. When the shift T is non-singular with respect to m, it will be called a *(generalised) Bernoulli shift* on X.

5.4. NOTATION.

For every integer n, let α_n be the partition of X that consists of the cylinder sets $[i]^n$, $i \in \{0,1,\dots,k-1\}$, then by A we shall denote the sub-σ-algebra generated by $\bigvee_{n=0}^{\infty} \alpha_n$. The σ-algebra A is called the *past σ-algebra*.

5.5. CALCULATION OF $J_m(A,T)$ FOR A GENERALISED MARKOV SHIFT T .

We assume, as in Definition 5.2, that the shift T is non-singular with respect to the Markov measure m constructed from the matrices P_n and row-vectors p_n satisfying conditions a) and b) of the same definition.

Let α be the state partition of X , i.e. the partition of X consisting of the cylinder sets $[i]^0$, $i \in \{0,1,\ldots,k-1\}$, then for a fixed but arbitrary positive integer ℓ , and a point $x = (x_n) \in X$, we have

$$I_m(\alpha|T^{-1}\alpha_0^{\ell})(x)$$

$$= - \log m_x^{\widehat{T^{-1}\alpha_0^{\ell}}}([x]_{\widehat{(\alpha_0^{\ell+1})}}) \quad ; \text{ by Definition 3.2}$$

$$= - \log \frac{m([x_0x_1\ldots x_\ell x_{\ell+1}]^0)}{m([x_1x_2\ldots x_\ell x_{\ell+1}]^1)}$$

$$= - \log \frac{p_0(x_0).P_0(x_0,x_1)}{p_1(x_1)}$$

; since m is a (generalised) Markov measure.

By letting $\ell \to \infty$, and using $I_m(\alpha|T^{-1}\alpha^-) = \lim_{\ell\to\infty} I_m(\alpha|T^{-1}\alpha_0^{\ell})$, (see [P1;Th.2.2]), we obtain

$$I_m(\alpha|T^{-1}\alpha^-)(x) = - \log \frac{p_0(x_0)P_0(x_0,x_1)}{p_1(x_1)} \quad .$$

Thus, since $T^{-1}A \subset A$,

$$K_m(A, T^{-1}A)(x) = I_m(A|T^{-1}A)(x)$$

$$= I_m(\alpha^-|T^{-1}\alpha^-)(x)$$

$$= I_m(\alpha|T^{-1}\alpha^-)(x)$$

$$= -\log \frac{p_0(x_0)P_0(x_0,x_1)}{p_1(x_1)} \quad . \tag{5.1}$$

For a fixed positive integer ℓ , we also have for m-a.e. $x = (x_n) \in B \in \alpha_0^\ell$,

$$E_m(\frac{dmT^{-1}}{dm}|\alpha_0^\ell)(x)$$

$$= \frac{1}{m(B)} \cdot \int_B \frac{dmT^{-1}}{dm}(x)dm(x)$$

$$= \frac{m(T^{-1}B)}{m(B)}$$

$$= \frac{p_1(x_0).P_1(x_0,x_1).\cdots.P_\ell(x_{\ell-1},x_\ell)}{p_0(x_0).P_0(x_0,x_1).\cdots.P_{\ell-1}(x_{\ell-1},x_\ell)} \quad .$$

Since T is non-singular with respect to m , the function dmT^{-1}/dm is m-integrable, and letting $\ell \to \infty$ we can apply the Increasing Martingale theorem to give

$$E_m(\frac{dmT^{-1}}{dm}|A)(Tx) = \frac{p_1(x_1)}{p_0(x_1)} . \prod_{i=1}^{\infty} (\frac{P_i(x_i,x_{i+1})}{P_{i-1}(x_i,x_{i+1})}) \quad . \tag{5.2}$$

By applying Definition 3.18, we get from equations (5.1) and (5.2) that for $x = (x_n) \in X$,

$J_m(A,T)(x)$

$$= - \log \frac{p_0(x_0)P_0(x_0,x_1)}{p_0(x_1)} \prod_{i=1}^{\infty} \left(\frac{P_i(x_i,x_{i+1})}{P_{i-1}(x_i,x_{i+1})}\right) . \qquad (5.3)$$

5.6. THE FORM OF $J_m(A,T)$ FOR A GENERALISED BERNOULLI SHIFT T .

We shall use the same notation as in Calculation 5.5. Since for the (generalised) Bernoulli shift the matrices P_n and row-vectors p_n satisfy condition c) of Definition 5.3, we may express equation (5.3) as:

$$J_m(A,T)(x) = - \log p_0(x_0) . \prod_{i=1}^{\infty} \left(\frac{p_i(x_i)}{p_{i-1}(x_i)}\right) . \qquad (5.4)$$

5.7. Remark.

On shift spaces, the concept of "finite co-ordinate change" or "finite dimensional homeomorphism" has proved very useful in providing invariants of various types of isomorphism between "topological Markov chains", see for example [C&K]and [Kri2]. We shall now calculate the form of the information cocycle, with respect to the past σ-algebra, for such transformations.

5.8. DEFINITION.

Let m be a generalised Markov measure on the shift space X with its product σ-algebra $\mathcal{B}$. A non-singular automorphism V of $(X,\mathcal{B},m)$ is called a *finite co-ordinate change* if there exists a positive integer ℓ such that for all $x = (x_n) \in X$,

$$x_i = (Vx)_i \ , \quad \text{for all} \quad i \in \mathbb{Z} \quad \text{with} \quad |i| \geq \ell \ . \qquad (5.5)$$

5.9. NOTATION.

Under composition, the countable set of all finite co-ordinate changes becomes a group which we shall denote by H .

As before we use T to denote the left-shift on $(X,\mathcal{B},m)$, and we assume that T is non-singular with respect to m . If A is the past σ-algebra then $H \subset \Gamma(A)$ (see Definition 3.17). Also, the group H is normalised by the shift T , i.e. $T^{-1}HT = H$, so that if we define $G: = \{T^kV; k \in \mathbb{Z}, V \in H\}$, then G is easily seen to be a countable subgroup of $\Gamma(A)$ that contains H .

We shall find it convenient to define one more subgroup of $\Gamma(A)$, as follows. Let H_- be the subgroup of H consisting of all those automorphisms V such that for all $x = (x_n) \in X$,

$$x_i = (Vx)_i \ , \quad \text{for all} \quad i \geq 0 \ .$$

5.10. THE FORM OF $J_m(A,V)$ FOR TRANSFORMATIONS $V \in H$.

For any $V \in H$, let $\ell = \ell(V)$ denote the smallest non-negative integer such that $T^\ell VT^{-\ell} \in H_-$, and to simplify some later expressions, let $S: = T^\ell VT^{-\ell}$.

Using equation (3.20) (the cocycle equation) from Theorem 3.19 we have

$$J_m(A,V) \equiv J_m(A,T^{-\ell}ST^{\ell})$$

$$= J_m(A,T^{\ell}) + J_m(A,S)\circ T^{\ell}$$

$$+ J_m(A,T^{-\ell})\circ ST^{\ell}$$

$$= J_m(A,T^{\ell}) + J_m(A,S)\circ T^{\ell}$$

$$- J_m(A,T^{\ell})\circ T^{-\ell}ST^{\ell} . \tag{5.6}$$

Since $S \in H_-$, $SA = A$, and thus

$$I_m(A|S^{-1}A) = I_m(S^{-1}A|A) = 0 \quad \text{m-a.e.}. \tag{5.7}$$

For each positive integer n we also have

$$E_m(\frac{dmS^{-1}}{dm}|\alpha_0^n) = 1 \quad \text{m-a.e.},$$

where α is the state partition of X , and hence by the Increasing Martingale theorem,

$$E_m(\frac{dmS^{-1}}{dm}|A) = 1 \quad \text{m-a.e.}. \tag{5.8}$$

As a consequence of equations (5.7) and (5.8), equation (5.6) reduces to

$$J_m(A,V) = J_m(A,T^{\ell}) - J_m(A,T^{\ell})\circ V . \tag{5.9}$$

We now use equation (5.3) and a straightforward computation to give

$$J_m(A,V) \equiv J_m(A,T^{-\ell}ST^{\ell})$$

$$= J_m(A,T^{\ell}) + J_m(A,S)\circ T^{\ell}$$

$$+ J_m(A,T^{-\ell})\circ ST^{\ell}$$

$$= J_m(A,T^{\ell}) + J_m(A,S)\circ T^{\ell}$$

$$- J_m(A,T^{\ell})\circ T^{-\ell}ST^{\ell} . \qquad (5.6)$$

Since $S \in H_-$, $SA = A$, and thus

$$I_m(A|S^{-1}A) = I_m(S^{-1}A|A) = 0 \quad \text{m-a.e..} \qquad (5.7)$$

For each positive integer n we also have

$$E_m\left(\frac{dmS^{-1}}{dm}\,|\alpha_0^n\right) = 1 \quad \text{m-a.e.,}$$

where α is the state partition of X , and hence by the Increasing Martingale theorem,

$$E_m\left(\frac{dmS^{-1}}{dm}\,|A\right) = 1 \quad \text{m-a.e..} \qquad (5.8)$$

As a consequence of equations (5.7) and (5.8), equation (5.6) reduces to

$$J_m(A,V) = J_m(A,T^{\ell}) - J_m(A,T^{\ell})\circ V . \qquad (5.9)$$

We now use equation (5.3) and a straightforward computation to give

$$J_m(A,V)(x) = \log \frac{p_0((Vx)_0)}{p_0(x_0)} \cdot \prod_{i=0}^{\infty} \left(\frac{P_i((Vx)_i,(Vx)_{i+1})}{P_i(x_i,x_{i+1})}\right) , \qquad (5.10)$$

where $x = (x_n) \in X$.

5.11. Remarks.

The infinite product appearing on the right-hand side of equation (5.10) makes sense because only finitely many of its terms take values that are different from unity.

Using Theorem 3.19, together with equations (5.10) and (5.3), it is possible to obtain explicit expressions for $J_m(A,V)$, when $V \in G$.

In the case of a shift space, with a shift invariant Markov probability measure, the corresponding groups H and G , and the form of the information cocycle for elements of these group, have recently been used to provide an invariant of finitary isomorphisms with finite expected coding lengths between such spaces. This invariant is due to W. Parry and K. Schmidt, and further details can be found in [B&S] and [P&S].

Within several settings, the so called 'information cocycle-coboundary equation' has been used to produce invariants of restrictive types of isomorphism. For an exposition of many of these results see [P&T2]. In the following subsection we shall describe some further invariants of the same information cocycle-coboundary equation.

INVARIANTS OF THE INFORMATION COCYCLE-COBOUNDARY EQUATION.

First, we define a particular type of isomorphism that we shall be considering. For a definition of a measure preserving Markov shift over countably many states, see Definitions 2.1.

5.12. DEFINITION.

Let (X_1,B_1,m_1,T_1), (X_2,B_2,m_2,T_2) be 2-sided, measure preserving Markov shifts over countably many states, then a measure preserving map ϕ between T_1 and T_2 satisfying $\phi T_1 = T_2\phi$ is called a *block code* if there exists a positive integer K such that for m_1-a.e. $x = (x_n) \in X_1$, and m_1-a.e. $y \in [x_{-K}x_{-K+1}\ldots x_K]^{-K}$, $(\phi(x))_0 = (\phi(y))_0$. If ϕ is also an isomorphism, then it is called a *block isomorphism*.

5.13. Remark and NOTATION.

Let X be a shift space over countably many states, and let m be a Markov probability measure defined on X (with its product σ-algebra B) by a stochastic matrix P, then there exists a shift invariant subspace $\bar{X}$ of X defined by

$$\bar{X} := \{x = (x_n) \in X;\ P(x_n,x_{n+1}) > 0 \text{ for all } n \in \mathbb{Z}\} .$$

The subspace $\bar{X}$ has full m-measure, i.e. $m(\bar{X}) = 1$, and if we endow the alphabet $A = \{0,1,2,\ldots\}$ from which X is constructed (see Definitions 2.1) with the discrete topology, and X with the corresponding product topology, then $\overline{X}$ is a closed subspace of X .

In the following we shall consider Markov shifts to be defined on this subspace, with the relative topology, and we shall use the (bar) notation $\overline{X}$ to indicate that we are in this situation.

Using the notation of Definition 5.12, we note that any block isomorphism $\phi:\overline{X}_1 \to \overline{X}_2$ coincides m_1-a.e. with a homeomorphism from $\overline{X}_1$ to $\overline{X}_2$.

5.14. NOTATION.

Let $(\overline{X},\mathcal{B},m,T)$ be a measure preserving Markov shift over a countable state space A , defined by a stochastic matrix P . For $n = 1,2,\ldots$, we use $\Delta_n = \Delta_n(P)$ to denote the ordered set consisting of all those positive numbers that arise from expressions of the form

$$P(x_0,x_1).P(x_1,x_2).\cdots.P(x_{n-1},x_0) ,$$

for $x_0x_1\ldots x_{n-1} \in \prod_{i=1}^{n} A$, where numbers occuring more than once are included in Δ_n as often as they occur, and the elements of Δ_n are arranged in descending order.

5.15. THEOREM.

The sequence of ordered sets $(\Delta_n)_{n\geq 1}$ is an invariant of block isomorphisms between measure preserving Markov shifts over countably many states.

Proof.

For $i = 1,2$, let $(\overline{X}_i, \mathcal{B}_i, m_i, T_i)$ be a measure preserving Markov shift over the countable alphabet A_i, defined by the stochastic matrix P_i, with past σ-algebra $\mathcal{A}_i$ (see Notation 5.4).

Let $\phi:\overline{X}_1 \to \overline{X}_2$ be a block isomorphism, then there exists an integer K such that $\phi^{-1}\mathcal{A}_2 \subset T_1^K\mathcal{A}_1$, and $\mathcal{A}_1 \subset T_1^K(\phi^{-1}\mathcal{A}_2)$. Using the fact that $\mathcal{A}_1 \underset{\sim}{\gamma} T_1^{-1}\mathcal{A}_1$, we can therefore apply Theorem 4.13, and utilise equations (3.21) and (3.22) of Theorem 3.20 to obtain the following equations.

$$
\begin{aligned}
&J_{m_1}(\mathcal{A}_1,T_1) \\
&= J_{m_1}(\phi^{-1}\mathcal{A}_2,T_1) + K_{m_1}(\mathcal{A}_1,\phi^{-1}\mathcal{A}_2) \\
&\quad - K_{m_1}(\mathcal{A}_1,\phi^{-1}\mathcal{A}_2)\circ T_1 \\
&= J_{m_2}(\mathcal{A}_2,T_2)\circ\phi + K_{m_1}(\mathcal{A}_1,\phi^{-1}\mathcal{A}_2) \\
&\quad - K_{m_1}(\mathcal{A}_1,\phi^{-1}\mathcal{A}_2)\circ T_1 \quad , \qquad (5.11)
\end{aligned}
$$

where the final equality holds because $\phi T_1 = T_2\phi$, $m_1\phi^{-1} = m_2$, and (cf. Lemma 3.8)

$$I_{m_1}(\phi^{-1}\mathcal{A}_2|T_1^{-1}\phi^{-1}\mathcal{A}_2) = I_{m_1\phi^{-1}}(\mathcal{A}_2|T_2^{-1}\mathcal{A}_2)\circ\phi \quad ,$$

$$I(m_1T_1^{-1},m_1|\phi^{-1}\mathcal{A}_2)\circ T_1 = I(m_2T_2^{-1},m_2|\mathcal{A}_2)\circ T_2\phi \quad .$$

Since ϕ is a block isomorphism the function $J_{m_1}(\mathcal{A}_1,T_1) - J_{m_2}(\mathcal{A}_2,T_2)\circ\phi$ depends on only finitely many co-ordinates.

From [P&T2;Corollary 2.42] (which although only presented for Markov shifts over finitely many states is nevertheless valid in the countable state case), this implies that the function $K_{m_1}(A_1, \phi^{-1}A_2)$ only depends on finitely many co-ordinates. By simultaneously altering both equation (5.11) and the map ϕ on a set of measure zero, we may therefore assume that equation (5.11) holds everywhere on $\bar{X}_1$, that all the functions appearing in it are continuous, and that the map ϕ (see Remark and Notation 5.13) is a homeomorphism. In the following we shall insist that equation (5.11) and the map ϕ have been adjusted in this manner.

Let $n \geq 1$ be an integer, and consider points $x \in \bar{X}_1$ that are fixed by the map T_1^n. Since ϕ (after adjustment) is a homeomorphism, it sends T_1^n-fixed points to T_2^n-fixed points, in a 1-1 way. Choose an arbitrary T_1^n-fixed point $x = (x_\ell) \in \bar{X}_1$, and let $y = (y_\ell) = \phi(x)$.

We now define measures on both $\bar{X}_1$ and $\bar{X}_2$ by placing mass $1/n$ on each point in the orbits of x and y. With respect to these measures the homeomorphism ϕ is measure preserving. By integrating both sides of the (adjusted) equation (5.11) with respect to this shift invariant measure, multiplying each side of the resulting equation by $-n$, and then exponentiating, we obtain:

$$P_1(x_0,x_1).P_1(x_1,x_2).\cdots.P_1(x_{n-1},x_0)$$

$$= P_2(y_0,y_1).P_2(y_1,y_2).\cdots.P_2(y_{n-1},y_0) .$$

Since the value of the expression on the left-hand side of this equation appears in $\Delta_n(P_1)$, and that on the right-hand side appears in $\Delta_n(P_2)$, by repeating the above argument for all T_1^n-fixed points in $\bar{X}_1$, and for all $n = 1,2,\ldots$, the result is proved.

□

5.16. <u>COROLLARY</u>. (Cf. [T1; Theorem 6] and [JKKM&S; Theorem 1])

Let $(X_1,\mathcal{B}_1,p,T_1)$, $(X_2,\mathcal{B}_2,q,T_2)$ be measure preserving Bernoulli shifts over countably many states, then they are block isomorphic precisely when the probability vectors that define the measures p and q are permutations of each other.

Proof.

The result follows from Theorem 5.15 by noting that for Bernoulli shifts over countably many states, the ordered set Δ_1 is a permutation of the defining probability vector.

□

5.17. <u>Remarks</u>.

The above result was originally stated for block isomorphisms between Bernoulli shifts over a finite number of states. Two independent proofs of this finite state Bernoulli case appear in [T1] and [JKKM&S].

One convenient form in which to display the ordered sets Δ_n , $n = 1,2,\ldots$, is as a zeta function. There are several ways to do this, but the following zeta function has been proved useful by W. Parry and S. Tuncel (see [P&T1]).

appears in $\Delta_n(P_1)$, and that on the right-hand side appears in $\Delta_n(P_2)$, by repeating the above argument for all T_1^n-fixed points in $\bar{X}_1$, and for all $n = 1,2,\ldots$, the result is proved.

□

5.16. <u>COROLLARY</u>. (Cf. [T1; Theorem 6] and [JKKM&S; Theorem 1])

Let $(X_1,\mathcal{B}_1,p,T_1)$, $(X_2,\mathcal{B}_2,q,T_2)$ be measure preserving Bernoulli shifts over countably many states, then they are block isomorphic precisely when the probability vectors that define the measures p and q are permutations of each other.

Proof.

The result follows from Theorem 5.15 by noting that for Bernoulli shifts over countably many states, the ordered set Δ_1 is a permutation of the defining probability vector.

□

5.17. <u>Remarks</u>.

The above result was originally stated for block isomorphisms between Bernoulli shifts over a finite number of states. Two independent proofs of this finite state Bernoulli case appear in [T1] and [JKKM&S].

One convenient form in which to display the ordered sets Δ_n, $n = 1,2,\ldots$, is as a zeta function. There are several ways to do this, but the following zeta function has been proved useful by W. Parry and S. Tuncel (see [P&T1]).

5.18. DEFINITION.

With the same assumptions as in Notation 5.14, for $i = 1,2$, let $r_i := \mathrm{card}(\Delta_i(P))$, where we allow the possibility $r_i = \infty$. We now label the (possibly infinite number of) values that occur in $\Delta_i(P)$ by listing them as

$$a_{i,1} \geq a_{i,2} \geq \cdots ,$$

where all values occuring more than once in $\Delta_i(P)$ are displayed the number of times they appear. (Note that the above sequence of inequalities may terminate.) We define the zeta function associated to the matrix P , to be the formal power series

$$\zeta(s,t) = \zeta(s,t)_P := \exp \sum_{i=1}^{\infty} \frac{s^i}{i} \sum_{j=1}^{r_i} (a_{i,j})^t . \tag{5.12}$$

5.19. Remarks.

For Markov shifts over finitely many states, the same zeta function (although defined in a slightly different manner) has been used by W. Parry and S. Tuncel in [P&T1]. In particular they show that it has (strictly) positive radius of convergence, and hence that the zeta function can be considered as a power series.

Some of the properties of the zeta function (for Markov shifts over finitely many states) that follow straightforwardly from its definition and Theorem 5.15, are listed below.

5.20. PROPERTIES OF $\zeta(s,t)$.

Let (X_1,B_1,m_1,T_1), (X_2,B_2,m_2,T_2) be block isomorphic Markov shifts over finitely many states, defined by stochastic matrices P and Q respectively, then:

a) $\zeta(s,t)_P = \zeta(s,t)_Q$

(If we consider this equality as being the equality of formal power series, it is also valid for block isomorphic Markov shifts over countably many states.)

b) $\zeta(s,1)_P$ = stochastic zeta function of P

$$= \frac{1}{\det(I-sP)} \quad ; \text{ see [P\&W] and [B\&L].}$$

c) $$\zeta(s,t)_P = \frac{1}{\det(I-s[P]^t)} ,$$

where for $t \in \mathbb{R}$, $[P]^t$ denotes the matrix whose $(i,j)^{th}$ entry is $(P(i,j))^t$ when $P(i,j) \neq 0$, and equals zero otherwise.

d) If P and Q are primitive matrices (that is, there exist positive integers k,ℓ such that $P^k > 0$ and $Q^\ell > 0$), then defining the numbers θ_i, $i \geq 1$, via the equation

$$\theta_i(t) := \sum_{j=1}^{r_i} (a_{i,j})^t ,$$

(with the notation of Definition 5.18) we have for any fixed $t \in \mathbb{R}$, that

$$\lim_{i\to\infty} \frac{\theta_{i+1}(t)}{\theta_i(t)}$$

$=$ maximum eigenvalue of $[P]^t$; see [Se1; Theorem 1.2]

$= \beta(t)$; ref. [T1],

and that this is an invariant of block isomorphism.

5.21. Remark.

The invariant $\beta(t)$ was found by S. Tuncel (see [T1]) and is an invariant of a wider class of isomorphisms than block isomorphisms. For further details we refer to [P&T3].

In order to show that it is possible for two "different" Markov shifts to be block isomorphic, we present the following example due to S. Tuncel and myself.

5.22. Example.

For $i = 1,2$, let $(X_i,\mathcal{B}_i,m_i,T_i)$ be the measure preserving Markov shift defined from the alphabet $A_i = \{0,1,2\}$ by the matrix P_i, where

$$P_1 = \begin{pmatrix} q & p & 0 \\ p & 0 & q \\ q & 0 & p \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} q & p^2 & pq \\ 1 & 0 & 0 \\ 0 & q & p \end{pmatrix} .$$

For an alphabet A, and $j \in \mathbb{Z}$, $\ell \in \mathbb{N}$, $\ell \geq 1$, we denote by $[A]_\ell^j$ the collection of all cylinder sets, with positive measure, of the form $[i_0 i_1 \ldots i_{\ell-1}]^j$, where $i_0, i_1, \ldots, i_{\ell-1} \in A$.

We now define two maps $\phi_0 : [A_1]_3^{-1} \to [A_2]_1^0$ and $\phi_0' : [A_2]_3^{-1} \to [A_1]_1^0$, as follows:

$$\phi_0(C) := \begin{cases} [0]^0 & ; \text{ if } C = [00\cdot]^{-1} \text{ or } [01\cdot]^{-1} \\ [1]^0 & ; \text{ if } C = [10\cdot]^{-1} \text{ or } [20\cdot]^{-1} \\ [2]^0 & ; \text{ if } C = [12\cdot]^{-1} \text{ or } [22\cdot]^{-1} \end{cases} ,$$

$$\phi_0'(C) := \begin{cases} [0]^0 & ; \text{ if } C = [\cdot 00]^{-1} \text{ or } [\cdot 10]^{-1} \\ [1]^0 & ; \text{ if } C = [\cdot 01]^{-1} \text{ or } [\cdot 02]^{-1} \\ [2]^0 & ; \text{ if } C = [\cdot 21]^{-1} \text{ or } [\cdot 22]^{-1} \end{cases} .$$

By applying the shifts T_1 and T_2 , the maps ϕ_0 and ϕ_0' can be "extended" to longer cylinders, and then used to construct maps $\phi : X_1 \to X_2$ and $\phi' : X_2 \to X_1$, in a straightforward way. We shall now prove that ϕ is a measure preserving isomorphism between T_1 and T_2 , with ϕ' as its inverse.

First note that by construction both ϕ and ϕ' intertwine the shifts T_1 and T_2 . To show that ϕ is an isomorphism it is sufficient to prove that $\phi\phi'$ = identity on X_2 , and $\phi'\phi$ = identity on X_1 . We shall prove that $\phi'\phi$ = identity on X_1 by showing that for m_1-a.e. point $x = (x_\ell) \in X_1$, $(\phi'\phi(x))_0 = x_0$. Using powers of the shifts, this implies that $(\phi'\phi(x))_n = x_n$, for all $n \in \mathbb{Z}$, and hence that $\phi'\phi$ = identity on X_1 . To check this condition we look at all cylinders in $[A_1]_3^{-1}$, and calculate the zero co-ordinate of $\phi'\phi(x)$, for a typical point x in each cylinder. The results of this calculation may be displayed as follows.

A	B	C
$[000]^{-1}, [001]^{-1}$	$[00]^0$	$[0]^0$
$[010]^{-1}$	$[01]^0$	$[1]^0$
$[012]^{-1}$	$[02]^0$	$[1]^0$
$[100]^{-1}, [101]^{-1}$	$[10]^0$	$[0]^0$
$[120]^{-1}, [220]^{-1}$	$[21]^0$	$[2]^0$
$[122]^{-1}, [222]^{-1}$	$[22]^0$	$[2]^0$
$[200]^{-1}, [201]^{-1}$	$[10]^0$	$[0]^0$

Table (5.13)

In Table (5.13), A, B and C are related by $\phi'\phi(A) \subset \phi'(B) \subset C$, and since the middle symbol of each cylinder set A is the same as the symbol in the corresponding cylinder set C, this shows that $\phi'\phi$ = identity on X_1. By a similar method one can check that $\phi\phi'$ = identity on X_2.

To prove that ϕ is measure preserving we shall produce a continuous function g such that

$$J_{m_1}(A_1, T_1) = J_{m_2}(A_2, T_2)\circ\phi + g\circ T_1 - g , \tag{5.14}$$

(compare equation (5.11)) where A_1 and A_2 are the past σ-algebras associated to T_1 and T_2 respectively. A simple calculation shows that T_1 and T_2 have the same entropy, and the existence of a continuous function g satisfying equation (5.14) implies (see [P&T2; Proposition 4.10]) that ϕ is measure preserving. The function g is defined by

$$g(x) := \begin{cases} 1 & ; \text{ if } x \in [00]^0 \text{ or } [01]^0 \\ 1 + \log p & ; \text{ if } x \in [10]^0 \text{ or } [22]^0 \\ 1 + \log q & ; \text{ if } x \in [12]^0 \text{ or } [20]^0 , \end{cases}$$

and using a straightforward calculation it can be shown to satisfy equation (5.14), thus ϕ is measure preserving.

We have now shown that the Markov shifts T_1 and T_2 are block isomorphic. It should be noted that in this example the Markov shifts T_1 and T_2 are very closely related; in fact T_2 is the inverse of T_1 , as may be seen by using equation (2.7).

5.23. Remark.

In several situations, for instance when using block isomorphisms or "regular isomorphisms" between Markov shifts over finitely many states, all the functions that appear in the information cocycle-coboundary equation (an example of which is equation (5.14)) are essentially bounded. The following calculation produces two invariants of such a cocycle-coboundary equation.

5.24. CALCULATION.

For $i = 1,2$, let $(X_i,\mathcal{B}_i,m_i)$ be a Lebesgue probability space, T_i be a non-singular automorphism of $(X_i,\mathcal{B}_i,m_i)$ and let $A_i \subset \mathcal{B}_i$ be a sub-σ-algebra satisfying $T_i^{-1}A_i \underset{I}{\sim} A_i$. Suppose also that $\phi : X_1 \to X_2$ is a measure-class preserving isomorphism satisfying $\phi T_1 = T_2\phi$, and that the cocycle-coboundary equation,

$$J_{m_1}(A_1,T_1) = J_{m_2}(A_2,T_2)\circ\phi + h\circ T_1 - h \,, \tag{5.15}$$

holds, where all the functions appearing are in $L^\infty(X_1,\mathcal{B}_1,m_1)$. Then the following calculation is valid.

$$\limsup_{n\to\infty} \frac{1}{n}(m_1\text{-ess.sup } J_{m_1}(A_1,T_1^n))$$

$$= \limsup_{n\to\infty} \frac{1}{n} (m_1\text{-ess.sup}(J_{m_2}(A_2,T_2^n)\circ\phi + h\circ T_1^n - h))$$

; using equation (5.15)

$$= \limsup_{n\to\infty} \frac{1}{n} (m_1\text{-ess.sup } J_{m_2}(A_2,T_2^n)\circ\phi)$$

$$= \limsup_{n\to\infty} \frac{1}{n} (m_2\text{-ess.sup } J_{m_2}(A_2,T_2^n))$$

; since ϕ preserves measure-class.

Similarly,

$$\liminf_{n\to\infty} \frac{1}{n} (m_1\text{-ess. inf } J_{m_1}(A_1, T_1^n))$$

$$= \liminf_{n\to\infty} \frac{1}{n} (m_2\text{-ess. inf } J_{m_2}(A_2, T_2^n)) .$$

If we define

$$h^+(J_{m_1}(A_1,T_1)) := \limsup_{n\to\infty} \frac{1}{n}(m_1\text{-ess. sup } J_{m_1}(A_1,T_1^n))$$

and

$$h^-(J_{m_1}(A_1,T_1)) := \liminf_{n\to\infty} \frac{1}{n}(m_1\text{-ess. inf } J_{m_1}(A_1,T_1^n)) ,$$

then the above calculation shows that $h^+(J_{m_1}(A_1,T_1))$ and $h^-(J_{m_1}(A_1,T_1))$ are invariants of equation (5.15).

The following theorem enables one to calculate the invariant $h^+(J_m(A,T))$ for Markov shifts over finitely many states. A similar result holds for $h^-(J_m(A,T))$.

5.25. THEOREM.

Let (X,B,m,T) be a measure preserving Markov shift over finitely many states, defined by a stochastic matrix $P = P(i,j)$, with past σ-algebra A , then

$$h^+(J_m(A,T))$$

$$= \max\{-\frac{1}{n} \log (P(x_0,x_1).P(x_1,x_2).\cdots.P(x_{n-1},x_0)) \ ;$$

$$n = 1,2,\ldots \text{ , the symbols } x_0,x_1,\ldots,x_{n-1}$$
$$\text{are pairwise distinct, and } m([x_0x_1\ldots x_{n-1}x_0]^0) > 0 \} \ .$$

Proof.

First note that since T is a measure preserving Markov shift, we can conclude from equation (5.3) that for $x = (x_n) \in X$,

$$J_m(A,T)(x) = -\log P(x_0,x_1) + \log p(x_1) - \log p(x_0) \ ,$$

where p is the (unique) strictly positive probability vector satisfying $p.P = p$ (see Definitions 2.1). When calculating $h^+(J_m(A,T))$ we therefore only need to consider the "adjusted" information cocycle defined by

$$J_m'(A,T)(x) := -\log P(x_0,x_1) \ .$$

To enable us to work almost entirely with strings of symbols (in preference to cylinder sets), we shall use the 1-1 correspondence defined by associating each cylinder set $[x_0x_1\ldots x_n]^0$ with the string $x_0x_1\ldots x_n$.

Let $A = \{0,1,\ldots,k-1\}$ be the alphabet from which the shift space X was constructed. For the purposes of this proof we shall make the following temporary definitions.

(i) A string $x_0x_1\ldots x_n$ of symbols from the alphabet A is called *admissible* if $m([x_0x_1\ldots x_n]^0) > 0$.

(ii) An admissible string of the form $x_0x_1\ldots x_nx_0$ is called a *cycle*.

(iii) A cycle $x_0x_1\ldots x_nx_0$ is called *prime* if $x_0,x_1,\ldots,x_n$ are pairwise distinct.

(iv) For an admissible string $s = x_0x_1\ldots x_n$, we define

$$h(s) := -\frac{1}{n}\log(P(x_0,x_1).P(x_1,x_2).\cdots.P(x_{n-1},x_n)) ,$$

$$\ell(s) := n ,$$

and we call $\ell(s)$ the *length* of s . If the string x_nx_0 is admissible then we can construct a cycle s' from s , by putting

$$s' := x_0x_1\ldots x_nx_0 \quad .$$

For such a pair of strings s and s' , we define

$$h'(s) := h(s') ,$$

$$\ell'(s) := \ell(s') .$$

Since there are only finitely many prime cycles, we can also define

$$h_{max} := \max\{h(s) \; ; \; s \text{ is a prime cycle}\} \; .$$

Let s' be a prime cycle such that $h(s') = h_{max}$, and define s^* to be the unique string such that $(s^*)' = s'$, then $h'(s^*) = h(s') = h_{max}$.

The following two facts will be useful:

(1) Any admissible string consisting of at least $k+1$ symbols contains a repeated symbol, and hence contains at least one prime cycle.

(2) There is a smallest strictly positive entry (denoted by P_{min}) in the transition matrix P .

Let us define

$$H := \limsup_{n\to\infty} \frac{1}{n} \left(m\text{-ess. sup } J_m(A,T^n)\right) ,$$

then H exists and is finite, because T is a Markov shift over only finitely many symbols.

For a fixed but arbitrary $\varepsilon > 0$, choose a corresponding integer $K > k$ such that

$$-\frac{1}{K}(\log P_{min}).(\ell'(s^*) + k+1) < \frac{\varepsilon}{2} \; . \tag{5.16}$$

and for the same ε , choose (and fix) an admissible string $s = x_0 x_1 \ldots x_u$ of length $\ell(s) = u > K$, for which

$$|h(s) - H| < \frac{\varepsilon}{4} \quad . \tag{5.17}$$

On the cylinder $[x_0 x_1 \ldots x_u]^0$ associated to the string s , the adjusted information cocycle satisfies

$$\frac{1}{u} J_m'(A, T^u) = h(s) \ ,$$

and the idea of the following proof is to show that by using the string s^* , it is possible to construct a cylinder of the same length on which

$$\frac{1}{u} J_m'(A, T^u) > h(s) - \frac{\varepsilon}{2} \ .$$

The arbitrary nature of the choice of ε , and the subsequent choice of the string s , will then imply the result.

Since $\ell(s) > K > k$, fact 1 implies that the string s contains a prime cycle. Let s_i' be the prime cycle that occurs closest to the left in the string s . If $s_i' = x_i x_{i+1} \ldots x_{i+j} x_{i+j+1}$, where $x_i = x_{i+j+1}$, and

$$s = x_0 x_1 \ldots x_{i-1} x_i x_{i+1} \ldots x_{i+j} x_{i+j+1} x_{i+j+2} \ldots x_u \ ,$$

then we shall denote by s_1 the string $x_i x_{i+1} \cdots x_{i+j}$ and by $s - s_1$ the admissible string

$$x_0 x_1 \cdots x_{i-1} x_{i+j+1} x_{i+j+2} \cdots x_u \ .$$

By removing in this manner the left-most prime cycle at each stage, we can continue this process and obtain a finite set of strings $s_1, s_2, \ldots, s_t$ (each constructed from a prime cycle by removing the right-most symbol) and a string $s_{res} := s - s_1 - s_2 - \ldots - s_t$, that contains no prime cycles. (Note that the above operations on strings are to be carried out in the order determined by reading the expression from left to right.) From fact 1 the length of s_{res} must be less than $k+1$, and as a consequence of the above definitions we also have

$$h(s).\ell(s) = h(s_{res}).\ell(s_{res}) + h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t) \ . \tag{5.18}$$

Using the definition of the cycle s^*, we have

$$h'(s^*)(\ell'(s_1) + \ldots + \ell'(s_t)) \geq h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t) \ , \tag{5.19}$$

and if r is the largest integer for which

$$r.\ell'(s^*) \leq \ell'(s_1) + \ldots + \ell'(s_t) \ ,$$

then

$$r.\ell'(s^*) \geq \ell'(s_1) + \ldots + \ell'(s_t) - \ell'(s^*) \ .$$

Combining this last inequality with inequality (5.19) gives

$$h'(s^*).r.\ell'(s^*) \geq h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t) - h'(s^*).\ell'(s^*) . \tag{5.20}$$

For all $j \in \mathbb{N}$, we shall denote by $\pi_j(s^*_\infty)$, the string consisting of the first j symbols of the infinite string s^*_∞ which is obtained by concatenating s^* with itself infinitely often.

Noting that $\ell(s) = u$, we now compare the values of $h(s)$ and $h(\pi_u(s^*_\infty))$ as follows:

$$h(s).u = h(s_{res}).\ell(s_{res}) + h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t)$$

; from equation (5.18)

$$\leq (k+1)(-\log P_{min}) + h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t) , \tag{5.21}$$

and

$$h(\pi_u(s^*_\infty)).u \geq h'(s^*).r.\ell'(s^*)$$

$$\geq h'(s_1).\ell'(s_1) + \ldots + h'(s_t).\ell'(s_t) - h'(s^*).\ell'(s^*)$$

; by inequality (5.20)

$$\geq h(s).u - (k+1).(-\log P_{min}) - \ell'(s^*)(-\log P_{min}) \tag{5.22}$$

; using inequality (5.21), and noticing that $h'(s^*) \leq (-\log P_{min})$.

Inequality (5.22) implies that

$$h(\pi_u(s^*_\infty)) \geq h(s) - \frac{1}{u}.(-\log P_{min})(k+1+\ell'(s^*))$$

$$\geq H - \frac{\varepsilon}{4} - \frac{\varepsilon}{2} \quad ; \text{ from inequalities (5.16) and (5.17)}$$

$$> H - \varepsilon \quad .$$

Since the choice of $\varepsilon > 0$ was arbitrary,

$$h'(s^*) = \lim_{u\to\infty} h(\pi_u(s^*_\infty))$$

$$= \limsup_{u\to\infty} h(\pi_u(s^*_\infty))$$

$$= H$$

$$= \limsup_{n\to\infty} \frac{1}{n} (m\text{-ess. sup } J_m(A,T^n)) \ ,$$

and this completes the proof.

□

5.26. Remark.

A similar proof to that of Theorem 5.25 shows that an analogous result holds for $h^-(J_m(A,T))$.

In the following proposition we show that for Bernoulli shifts over finitely many states there is a connection between the two invariants $h^+(J_m(A,T))$ and $h^-(J_m(A,T))$, and the function $\beta(t)$

referred to in Remark 5.21. (For further details of $\beta(t)$ see [T1] and [P&T3].) We conjecture that this relation is also valid for Markov shifts over finitely many states.

5.27. PROPOSITION.

Let $(X,\mathcal{B},m,T)$ be a measure preserving Bernoulli shift over finitely many states, defined by a stochastic matrix P, and with past σ-algebra A. For all $t \in \mathbb{R}$, define $\beta(t) := $ maximum eigenvalue of $[P]^t$, where $[P]^t$ denotes the matrix whose $(i,j)^{th}$ entry is $(P(i,j))^t$. Then

$$h^-(J_m(A,T)) = \log\left(\lim_{t\to\infty} \frac{\beta(t)}{\beta(t+1)}\right) , \tag{5.23}$$

and

$$h^+(J_m(A,T)) = \log\left(\lim_{t\to-\infty} \frac{\beta(t)}{\beta(t+1)}\right) . \tag{5.24}$$

Proof.

Let $p = (p(0),p(1),\cdots,p(k-1))$ be the (unique) strictly positive probability vector satisfying $p.P = p$, then (using a result from [T1])

$$\beta(t) = \sum_{i=0}^{k-1} (p(i))^t ,$$

and the following calculation therefore proves equation (5.24).

$$\log\left(\lim_{t\to-\infty} \frac{\beta(t)}{\beta(t+1)}\right)$$

$$= \log\left(\lim_{t\to-\infty} \left(\sum_{i=0}^{k-1} (p(i))^t\right)/\left(\sum_{i=0}^{k-1} (p(i))^{t+1}\right)\right)$$

$$= -\log \min\{p(0), p(1), \ldots, p(k-1)\}$$

$$= h^+(J_m(A,T)) \;,$$

where the final equality follows from Theorem 5.25. Equation (5.23) can be proved by a similar method.

□

To conclude this section we shall now provide an example of two pairs of Bernoulli shifts that, although neither block isomorphic nor "regularly isomorphic" (as a consequence of Corollary 5.16 and [T1; Theorem 12]), are not distinguishable using presently known invariants of other types of isomorphism.

5.28. <u>Example</u>.

Consider the Bernoulli shifts constructed from the following probability vectors, where we use the notation $a*b$ to denote b copies of a .

$$p_1 = (\tfrac{1}{4}, \tfrac{1}{8} * 2, \tfrac{1}{16}, \tfrac{1}{32} * 13, \tfrac{1}{64} * 2) \;,$$

$$q_1 = (\tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16} * 7, \tfrac{1}{32}, \tfrac{1}{64} * 10) \;.$$

By straightforward calculations, these Bernoulli shifts can be shown to have the same entropy, information variance σ^2 (see [F&P]), and group invariant Λ (see [P6]). Invariants depending on groups generated by elements, or ratios of elements, from the vectors p_1 and q_1 (see [P&S], [B&S]) are also seen to be the same for both shifts.

It has recently been shown by W. Parry, that for Bernoulli shifts over finitely many states the 3^{rd}-moment of the information cocycle is an invariant of information cocycle-coboundary equations where all the functions are in L^3 . Such an equation holds between the information cocycles of "regularly isomorphic" Bernoulli shifts over finitely many states. The following pair of probability vectors define Bernoulli shifts that satisfy all the properties mentioned above for p_1 and q_1 , and they also have the same 3^{rd}-moment invariant.

$$p_2 = (\tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16} * 2, \tfrac{1}{32}, \tfrac{1}{64} * 25, \tfrac{1}{128}, \tfrac{1}{256} * 18) \quad ,$$

$$q_2 = (\tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16}, \tfrac{1}{32} * 9, \tfrac{1}{64}, \tfrac{1}{128} * 33, \tfrac{1}{256} * 2) \quad .$$

We remark that for any fixed integer $n \geq 1$, it is possible to construct Bernoulli shifts with all of the above properties and with equal i^{th}-moments, for all $i = 2,\dots,n$.

Section 6.

ERGODIC NON-SINGULAR BERNOULLI SHIFTS.

In [Kr], U. Krengel produced an example of a product probability measure (or Bernoulli measure) on a shift space over two states, with respect to which the shift was non-singular and ergodic. This measure was also not equivalent to any shift invariant probability measure. By altering his construction slightly, we are able to use an arbitrary positive probability vector (of length greater than one) to obtain an uncountable infinity of mutually singular product probability measures, all on the same shift space. With respect to each such measure, the shift is shown to be non-singular and ergodic, and to have a bounded Radon-Nikodym derivative. Also, none of these measures is equivalent to a shift invariant probability measure.

Since it was not known if the measure constructed by U. Krengel was equivalent to a shift invariant σ-finite measure, T. Hamachi [Ha] produced a product probability measure on the shift space over two states for which no equivalent σ-finite invariant measure exists, and with respect to which the shift is non-singular. By slightly modifying T. Hamachi's basic method, we produce an uncountable infinity of such measures, all mutually singular, and all defined on the same shift space. The calculations used by T. Hamachi to show that his measure satisfies the required properties are lengthy, and in order to avoid duplicating them, we provide a construction that allows us to use his work for each of the measures we obtain. A complete copy of T. Hamachi's paper is included as an appendix.

6.1. Remark.

In the following construction we shall use a probability vector of length greater than one, to produce an uncountable infinity of product probability measures p^λ , $0 \le \lambda \le 1$, on an associated shift space. The measures p^λ , $0 \le \lambda \le 1$, all satisfy:

a) The shift T is non-singular and ergodic with respect to p^λ .

b) The Radon-Nikodym derivative $dp^\lambda T/dp^\lambda$ is bounded away from zero and infinity.

c) There exists no invariant probability measure $\mu \ll p^\lambda$.

d) If $0 \le \lambda_1 < \lambda_2 \le 1$, then $p^{\lambda_1} \perp p^{\lambda_2}$.

6.2. CONSTRUCTION. (Ref. [Kr])

Choose and fix a strictly positive probability vector p_0 , where $p_0 = (p(1),p(2),\dots,p(u))$, $u \ge 2$, and $p(1) + \dots + p(u) = 1$. Let λ be a parameter, $0 \le \lambda \le 1$, then we shall construct a product probability measure p^λ on the infinite product space

$$X = \prod_{i=-\infty}^{\infty} \{1,2,\dots,u\} ,$$

with the σ-algebra generated by cylinders. For each λ , $0 \le \lambda \le 1$, we construct a sequence of probability measures $\{p_i^\lambda\}_{i=-\infty}^{\infty}$ on the set $\{1,2,\dots,u\}$, and define

$$p^{\lambda} := \prod_{i=-\infty}^{\infty} p_i^{\lambda} \ .$$

For $i \geq 0$, we define $p_i^{\lambda} := p_0$, the construction of p_i^{λ}, for $i < 0$ is carried out below.

Choose and fix a real number d satisfying

$$0 < d < \min\{p(1), p(2), \ldots, p(u)\} \tag{6.1}$$

and a decreasing sequence of real numbers $\{\delta_i\}$ satisfying both

$$\delta_i \geq \delta_{i+1} > 1 \ , \ \text{for all } i = 1,2,\ldots \ , \tag{6.2}$$

and

$$\prod_{i=1}^{\infty} \delta_i < 2 \ . \tag{6.3}$$

For $j \in \mathbb{N}$ we shall inductively define natural numbers m_j, n_j, k_j and ℓ_j such that

$$n_j = m_j + k_j \quad , \tag{6.4}$$

$$m_{j+1} = n_j + \ell_j \ , \tag{6.5}$$

and real numbers ϵ_j ; $\epsilon_j^{\lambda}(i)$, $i = 1,2,\ldots,u$; $0 \leq \lambda \leq 1$. From these numbers we define the measures p_i^{λ}, $i < 0$, $0 \leq \lambda \leq 1$, by setting $p_i^{\lambda} := (p_i^{\lambda}(1), p_i^{\lambda}(2), \ldots, p_i^{\lambda}(u))$, where

$$p_i^\lambda(s) := \begin{cases} p_0(s) - \varepsilon_j^\lambda(s) \; ; \; i = -m_j, -m_j-1, \ldots, -n_j+1 \\ p_0(s) \; ; \; i = -n_j, -n_j-1, \ldots, -m_{j+1}+1 \end{cases} \quad (6.6)$$

for $s = 1,2,\ldots,u$; $0 \le \lambda \le 1$.

To simplify the construction, let us define $\varepsilon_j^\lambda(s) := 0$, for all $j \in \mathbb{N}$, $s = 3,4,\ldots,u$; and all $0 \le \lambda \le 1$. We shall therefore only alter the values of $\varepsilon_j^\lambda(1)$ and $\varepsilon_j^\lambda(2)$, for $j \in \mathbb{N}$ and $0 \le \lambda \le 1$. It is possible to produce a construction that varies all of the $\varepsilon_j^\lambda(s)$, for $j \in \mathbb{N}$, all $s = 1,2,\ldots,u$; and all $0 \le \lambda \le 1$. However the construction is necessarily more complicated, and adds little to the present result.

The numbers $\varepsilon_j^\lambda(1)$, $\varepsilon_j^\lambda(2)$, $j \in \mathbb{N}$, $0 \le \lambda \le 1$ remain to be defined.

We start the induction by setting $\varepsilon_1 := d/2$, $m_1 := 1$, and $k_1 := 1$. From equation (6.4), $n_1 = m_1+k_1 = 2$. Define $\varepsilon_1^\lambda(i)$, $i = 1,2$; $0 \le \lambda \le 1$ by

$$\varepsilon_1^\lambda(1) := (1+\lambda).\varepsilon_1/2 \quad , \quad (6.7)$$

$$\varepsilon_1^\lambda(2) := -\varepsilon_1^\lambda(1) \quad . \quad (6.8)$$

Using the defining equation (6.6), we have therefore defined p_{-1}^λ and p_{-2}^λ , for all $0 \le \lambda \le 1$. Now define

$$p_i^{\lambda,(1)} := \begin{cases} p_i^{\lambda} & ; \ i \geq -2 \\ p_0 & ; \ i < -2 \end{cases} \tag{6.9}$$

and let

$$\beta_1 := \inf_{0 \leq \lambda \leq 1} \quad \inf_{\substack{n,k \in \mathbb{N} \\ \{x_{-n},\ldots,x_0\} \in \{1,\ldots,u\}^{n+1}}} \left\{ \prod_{i=-n}^{0} \left(\frac{p_{i+k}^{\lambda,(1)}(x_i)}{p_i^{\lambda,(1)}(x_i)} \right) \right\}. \tag{6.10}$$

The number β_1 is positive, since at most two of the factors on the right-hand side of equation (6.10) can be different from 1 , and the range of values of the remaining factors is bounded uniformly away from zero, for all $0 \leq \lambda \leq 1$, simultaneously.

Since $\beta_1 > 0$, we can find an $r_1 \in \mathbb{N}$ for which

$$r_1 \beta_1 > 2 \quad , \tag{6.11}$$

and from such an r_1 we define

$$\ell_1 := 5r_1 \quad . \tag{6.12}$$

The first step of the inductive construction is complete.

At the end of step t of the construction, we have defined natural numbers

$$m_1 < n_1 < m_2 < \dots < m_t < n_t \quad , \tag{6.13}$$

$$\ell_1 < \ell_2 < \dots < \ell_t \quad , \tag{6.14}$$

$$k_1 < k_2 < \dots < k_t \quad , \tag{6.15}$$

satisfying equations (6.4) and (6.5), and we have also chosen real numbers

$$\varepsilon_1 > \varepsilon_2 > \dots > \varepsilon_t > 0 \quad , \tag{6.16}$$

$$\varepsilon_j^\lambda(i) \; ; \; i = 1,2,\dots,u \; ; \; j = 1,\dots,t \; ; \; 0 \le \lambda \le 1 \quad , \tag{6.17}$$

together with natural numbers

$$r_1, r_2, \dots, r_t \quad . \tag{6.18}$$

At the beginning of step $(t+1)$, m_{t+1} is defined by equation (6.5). We now choose an $\varepsilon_{t+1} > 0$ for which all of the conditions

$$\varepsilon_{t+1} < \varepsilon_t/2 \quad , \tag{6.19}$$

$$\varepsilon_{t+1} < 2^{-(t+1)} \quad , \tag{6.20}$$

$$1 < \left(\frac{p_0(i) + \varepsilon_{t+1}}{p_0(i) - \varepsilon_{t+1}}\right)^{2m_{t+1}} < \delta_t \; ; \; i = 1,2 \quad , \tag{6.21}$$

are satisfied. Let us define $\varepsilon_{t+1}^\lambda(i) \; ; \; i = 1,2 \; ; \; 0 \le \lambda \le 1$ by

$$\varepsilon^{\lambda}_{t+1}(1) := (\tfrac{1+\lambda}{2}).\varepsilon_{t+1} , \tag{6.22}$$

$$\varepsilon^{\lambda}_{t+1}(2) := -\varepsilon^{\lambda}_{t+1}(1) , \tag{6.23}$$

and find a $k_{t+1} \in \mathbb{N}$ that satisfies

$$k_{t+1} > k_t , \tag{6.24}$$

$$k_{t+1}.(\varepsilon_{t+1})^2.2^{-2(t+1)} \geq 1 , \tag{6.25}$$

$$k_{t+1} > 5r_t . \tag{6.26}$$

From equation (6.4), $n_{t+1} = m_{t+1} + k_{t+1}$, and therefore p^{λ}_i has been defined for $i > -n_{t+1}$, $0 \leq \lambda \leq 1$. Now set

$$p^{\lambda,(t+1)}_i := \begin{cases} p^{\lambda}_i & ; \ i > -n_{t+1} \\ p_0 & ; \ i \leq -n_{t+1} \end{cases} \tag{6.27}$$

and define

$$\beta_{t+1} := \inf_{0\leq\lambda\leq 1} \ \inf_{\substack{n,k\in\mathbb{N} \\ (x_{-n},\ldots,x_0)\in\{1,\ldots,u\}^{n+1}}} \left\{ \prod_{i=-n}^{0} \left(\frac{p^{\lambda,(t+1)}_{i+k}(x_i)}{p^{\lambda,(t+1)}_i(x_i)} \right) \right\}. \tag{6.28}$$

The number β_{t+1} is positive because at most $2n_{t+1}$ of the factors on the right-hand side of equation (6.28) can be different from 1, and the range of values of the remaining factors is bounded uniformly away from zero, for all $0 \le \lambda \le 1$, simultaneously.

Next, we find an $r_{t+1} \in \mathbb{N}$ satisfying

$$r_{t+1} > 2r_t , \tag{6.29}$$

$$r_{t+1} \cdot \beta_{t+1} > 2 , \tag{6.30}$$

and define

$$\ell_{t+1} := 5r_{t+1} + n_t . \tag{6.31}$$

Step $(t+1)$ of the inductive construction is complete.

6.3. Remarks.

In order to show that the measures produced in Construction 6.2 satisfy the required properties, some general facts are needed.

Suppose p,q are two product measures defined on the infinite product space X, with its product σ-algebra. If

$$p = \prod_{i=-\infty}^{\infty} p_i \quad , \quad p_i = (p_i(1), p_i(2), \ldots, p_i(u)) ,$$

and

$$q = \prod_{i=-\infty}^{\infty} q_i \quad , \quad q_i = (q_i(1), q_i(2), \ldots, q_i(u)) ,$$

where $p_i \sim q_i$, for all $i \in \mathbb{Z}$, then by Kakutani's Theorem (see [K]), p and q are equivalent or singular according as the infinite series

$$\sum_{i=-\infty}^{\infty} 2(1 - \sum_{j=1}^{u} (p_i(j)q_i(j))^{\frac{1}{2}}) \tag{6.32}$$

is convergent or divergent. However,

$$2(1 - \sum_{j=1}^{u} (p_i(j)q_i(j))^{\frac{1}{2}}) = \sum_{j=1}^{u} (\sqrt{p_i(j)} - \sqrt{q_i(j)})^2 ,$$

and hence the series in (6.32) converges if and only if

$$\sum_{i=-\infty}^{\infty} (\sum_{j=1}^{u} (\sqrt{p_i(j)} - \sqrt{q_i(j)})^2)$$

converges.

We also have

$$\sum_{j=1}^{u} (\sqrt{p_i(j)}-\sqrt{q_i(j)})^2 = \sum_{j=1}^{u} p_i(j)(1 - (1 + \frac{q_i(j)-p_i(j)}{p_i(j)})^{\frac{1}{2}})^2 .$$

If there exists a real number $a_0 > 0$ such that

$$0 < a_0 \le p_i(j), q_i(j) \le 1-a_0 < 1; \ \forall\ j = 1,\dots,u; \ \forall\ i \in \mathbb{Z} , \tag{6.33}$$

then the series in (6.32) converges if and only if

$$\sum_{i=-\infty}^{\infty} \left(1 - \left(1 + \frac{q_i(j) - p_i(j)}{p_i(j)}\right)^{\frac{1}{2}}\right)^2 \tag{6.34}$$

converges.

The proof of the following lemma is straightforward and is omitted.

6.4. LEMMA.

Given any $a_0 > 0$, there exist constants $c_1, c_2 > 0$ such that

$$c_1 x^2 \le (1 - (1+x)^{\frac{1}{2}})^2 \le c_2 x^2 ,$$

for all x satisfying $|x| \le 1 - a_0$.

6.5. Remark.

If condition (6.33) is satisfied, then Lemma 6.4 implies that the convergence of the series in (6.32) is equivalent to the convergence of

$$\sum_{i=-\infty}^{\infty} \left(\sum_{j=1}^{u} (p_i(j) - q_i(j))^2 \right) , \tag{6.35}$$

which for such measures converges or diverges according as the measures p and q are equivalent or singular.

We now prove that Construction 6.2 produces measures that satisfy the conditions of Remark 6.1.

6.6. PROPOSITION.

Let p^λ, $0 \le \lambda \le 1$, be any one of the measures produced in Construction 6.2, then p^λ satisfies:

a) The shift T is non-singular and ergodic with respect to the measure p^λ.

b) The Radon-Nikodym derivative $dp^\lambda T/dp^\lambda$ is bounded away from zero and infinity.

c) There exists no invariant probability measure $\mu \ll p^\lambda$.

The measures p^λ, $0 \le \lambda \le 1$ also satisfy

d) If $0 \le \lambda_1 < \lambda_2 \le 1$, then $p^{\lambda_1} \perp p^{\lambda_2}$.

Proof.

The proof will be divided into four parts.

a) The shift transformation T is non-singular and ergodic with respect to p^λ

The measures p^λ, $0 \le \lambda \le 1$, all satisfy condition (6.33), and hence we may apply Remark 6.5. For a fixed λ, $0 \le \lambda \le 1$, the co-ordinate measures p_i^λ and p_{i+1}^λ only differ for numbers i of the form $i = -m_t$, or $i = -n_t$. In both of these cases $|p_i(j) - p_{i+1}(j)| \le \epsilon_t < 2^{-t}$, for all $j = 1,\dots,u$, and hence, using Remark 6.5, the following calculation proves that T is non-singular:

$$\sum_{i=-\infty}^{\infty} \left(\sum_{j=1}^{u} (p_i(j) - p_{i+1}(j))^2 \right)$$

$$= \sum_{i=-\infty}^{0} \left(\sum_{j=1}^{u} (p_i(j) - p_{i+1}(j))^2 \right)$$

$$\leq 2 \sum_{t=1}^{\infty} (\varepsilon_t)^2$$

$$\leq 2 \sum_{t=1}^{\infty} (2^{-t})^2 < \infty \quad .$$

We now prove that T is conservative with respect to p^λ . Let $x = (x_i)$ be a point in X , then for an arbitrary, but temporarily fixed, integer $k \geq 1$, we shall look at

$$\frac{dp^\lambda T^{-k}}{dp^\lambda}(x) \quad .$$

If t is such that

$$k \leq r_1 + r_2 + \ldots + r_t \tag{6.36}$$

$$\leq 2r_t \qquad ; \text{ by inequality (6.29),}$$

and if $i \geq -n_t - 2r_t$, then $p_i^\lambda = p_i^{\lambda,(t)}$, and $p_{i+k}^\lambda = p_{i+k}^{\lambda,(t)}$, by the defining equation (6.27). Using this and the definition of β_t (equation (6.30)), we have that for $n \leq n_t + 2r_t$,

$$\prod_{i=-n}^{0} \left(\frac{p_{i+k}^{\lambda}(x_i)}{p_i^{\lambda}(x_i)}\right) \geq \beta_t \quad . \tag{6.37}$$

If $i < -n_t - 2r_t$, the ratio $p_{i+k}^{\lambda}(x_i)/p_i^{\lambda}(x_i)$ can only be different from 1 , if i is in one of the disjoint sets I_s , $s > t$, defined by

$$I_s: = \{-m_s,-m_s-1,\ldots,-m_s-k+1,-n_s,-n_s-1,\ldots,-n_s-k+1\} \; . \tag{6.38}$$

If $i \in I_s$, then

$$\frac{p_0(x_i)-|\varepsilon_s(x_i)|}{p_0(x_i)+|\varepsilon_s(x_i)|} \leq \frac{p_{i+k}^{\lambda}(x_i)}{p_i^{\lambda}(x_i)} \leq \frac{p_0(x_i)+|\varepsilon_s(x_i)|}{p_0(x_i)-|\varepsilon_s(x_i)|} \quad . \tag{6.39}$$

Since

$$\begin{aligned} |I_s| &= 2k \\ &\leq 4r_t \quad ; \text{ by inequalities (6.29) and (6.36)} \\ &\leq \ell_t \quad ; \text{ using equation (6.31)} \\ &\leq 2m_s \quad ; \text{ from equation (6.5), because } s > t , \end{aligned}$$

therefore, for all $n = 1,2,\ldots$

$$\prod_{i=-n}^{0} \left(\frac{p_{i+k}^{\lambda}(x_i)}{p_i^{\lambda}(x_i)}\right)$$

$$\geq \beta_t \cdot \prod_{s=2}^{\infty}\left(\frac{1}{\delta_s}\right) \quad ; \text{ by inequalities (6.21) and (6.39)}$$

$$\geq \frac{\beta_t}{2} \quad ; \text{ using inequality (6.3).}$$

The above calculation proves that

$$\frac{dp^{\lambda}T^{-k}}{dp^{\lambda}}(x) \geq \frac{\beta_t}{2} \ ,$$

for p^{λ}-a.e. $x \in X$, and $k = r_1 + r_2 + \ldots + r_{t-1} + j$, where $j = 1,2,\ldots,r_t$. Collecting all of the above inequalities together, we obtain

$$\sum_{k=1}^{\infty} \left(\frac{dp^{\lambda}T^{-k}}{dp^{\lambda}}\right)(x) \geq \sum_{t=1}^{\infty} \frac{\beta_t \cdot r_t}{2} \quad p^{\lambda} \text{ - a.e.}$$

$$= \infty \quad ; \text{ from inequality (6.30).}$$

By a result of J. Choksi [Ch], this implies that the shift transformation T is conservative with respect to the measure p^{λ} .

If we now apply U. Krengel's result (see [Kr]) that for measures of the form p^{λ} , where for $i \geq 0$, $p_i^{\lambda} = p_0$, a fixed co-ordinate measure; conservative implies ergodic, then the result of part a) is proved.

b) <u>The Radon-Nikodym derivative $dp^\lambda T/dp^\lambda$ is bounded away from zero and infinity</u>.

For a fixed λ, $0 \le \lambda \le 1$, the co-ordinate measures p_i^λ and p_{i+1}^λ differ only for numbers of the form $i = -m_t$, or $i = -n_t$. For such measures we have, as a consequence of inequality (6.21), that

$$\frac{1}{\delta_{t-1}} < \frac{dp_{i+1}^\lambda}{dp_i^\lambda} < \delta_{t-1} ,$$

and hence using inequality (6.3)

$$\frac{1}{2} < \frac{dp^\lambda T}{dp^\lambda} < 2 ,$$

and this completes the proof of part b).

c) <u>There exists no T invariant probability measure $\mu \ll p^\lambda$</u> .

Denote by $\mathcal{P}$ the past σ-algebra consisting of the product σ-algebra generated by the non-negative co-ordinates, i.e. by the co-ordinate maps X_i, for $i \ge 0$. For a measure μ, we use the notation $(\mu|\mathcal{P})$ to represent the restriction of μ to $\mathcal{P}$.

Choose and fix an arbitrary λ, $0 \le \lambda \le 1$, and suppose μ is an invariant probability measure, $\mu \ll p^\lambda$, then $(\mu|\mathcal{P}) \ll (p^\lambda|\mathcal{P}) = (q|\mathcal{P})$, where the measure q is defined by

$$q := \prod_{i=-\infty}^{\infty} q_i , \quad q_i = p_0 \text{ for all } i \in \mathbb{Z} .$$

Since the map T on $(X,\mathcal{P},q)$ is the one-sided Bernoulli shift derived from the probability vector p_0, it is therefore ergodic. This implies that $(\mu|\mathcal{P}) = (q|\mathcal{P})$, and as μ is invariant, $\mu = q$. To show that $q \perp p^\lambda$ and hence complete the proof that there exists no T invariant probability measure $\mu \ll p^\lambda$, Remark 6.5 together with the following calculation is therefore sufficient.

$$\sum_{i=-\infty}^{\infty} \left(\sum_{j=1}^{u} (p_i^\lambda(j) - q_i(j))^2 \right)$$

$$\geq \sum_{t=2}^{\infty} k_t \cdot (\varepsilon_t^\lambda)^2$$; because there are precisely k_t values of i for which $p_i^\lambda(s) = p_0(s) - \varepsilon_t(s)$, $s = 1,2,\ldots,u$

$$= \infty$$; using inequality (6.25).

This completes the proof of part c).

d) <u>If $0 \leq \lambda_1 < \lambda_2 \leq 1$, then $p^{\lambda_1} \perp p^{\lambda_2}$</u>.

We use a similar technique to that used in part c). Suppose $0 \leq \lambda_1 < \lambda_2 \leq 1$, then for any integer $t \geq 1$, there exist k_t values of i for which

$$p_i^{\lambda_1}(s) = p_0(s) - \varepsilon_t^{\lambda_1}(s) \text{ , and}$$

$$p_i^{\lambda_2}(s) = p_0(s) - \varepsilon_t^{\lambda_2}(s) \text{ .}$$

for $s = 1,2,\ldots,u$.

Applying Remark 6.5, the following calculation proves that $p^{\lambda_1} \perp p^{\lambda_2}$, and hence completes the proof.

$$\sum_{i=-\infty}^{\infty} \left(\sum_{j=1}^{u} (p_i^{\lambda_1}(j) - p_i^{\lambda_2}(j))^2 \right)$$

$$\geq \sum_{t=2}^{\infty} k_t \left(\sum_{j=1}^{u} (\varepsilon_t^{\lambda_1}(j) - \varepsilon_t^{\lambda_2}(j))^2 \right)$$

$$\geq \sum_{t=2}^{\infty} k_t \cdot \varepsilon_t^2 \cdot \left(\frac{\lambda_2-\lambda_1}{2}\right)^2$$

$$= \infty \qquad ; \text{ by inequality (6.25).}$$

The proof of Propositicn 6.6 is now complete.

□

6.7. Remark.

In Construction 6.2 we produced an uncountable infinity of mutually singular probability measures p^λ , $0 \leq \lambda \leq 1$, on a shift space, with the property that for each measure p^λ there existed no shift invariant probability measure $\mu \ll p^\lambda$. It is not known if any of these measures is equivalent to an invariant σ-finite measure. However, T. Hamachi [Ha] has constructed a product measure for which no equivalent σ-finite invariant measure exists. By slightly modifying his basic method, we produce an uncountable infinity of such measures, all mutually singular, and all defined on the shift space over two symbols. In order to avoid duplicating T. Hamachi's calculations, we provide a construction that immediately allows us to use his work, for all of the measures. The measures are then shown to be mutually singular.

6.8. NOTATION.

Let X be the doubly infinite product space over the two symbols, 0 and 1 , i.e.

$$X = \prod_{i=-\infty}^{\infty} \{0,1\} \ ,$$

and let T be the left-shift on X . We shall produce a set $A \subset [0,1]$, of positive Lebesgue measure, together with a collection of product measures p^{α} , $0 \le \alpha \le 1$, on X , for which if $\alpha \in A$, then p^{α} satisfies all of the requirements of the measure p constructed by T. Hamachi in [Ha] , and hence the measure p^{α} , $\alpha \in A$, is not equivalent to any σ-finite shift invariant measure. Each of the measures $p^{\alpha} := \prod_{k=-\infty}^{\infty} p_k^{\alpha}$, $0 \le \alpha \le 1$, will be defined by

$$p_k^{\alpha} := \begin{cases} \mu & ; \text{ if } k \ge 0 \\ \nu_t^{\alpha} & ; \text{ if } -N_t < k \le -M_{t-1} \\ \mu & ; \text{ if } -M_t < k \le -N_t \end{cases} \qquad (6.40)$$

for $t \ge 1$, where μ is the measure on $\{0,1\}$ given by $\mu(0) = \mu(1) = \frac{1}{2}$. The measures ν_t^{α} are defined by setting

$$\nu_t^\alpha(s) := \begin{cases} 1/(1+\lambda_t^\alpha) & ; \quad s = 0 \\ \lambda_t^\alpha/(1+\lambda_t^\alpha) & ; \quad s = 1 \end{cases} \tag{6.41}$$

where the λ_t^α are real numbers yet to be defined, and M_t , N_t , n_t and m_t are positive integers that satisfy:

$$N_t := M_{t-1} + n_t \quad , \tag{6.42}$$

$$M_t := N_t + m_t \quad , \tag{6.43}$$

$$M_0 := 1 \quad . \tag{6.44}$$

To obtain the set $A \subset [0,1]$, and the collection of measures p^α , $\alpha \in [0,1]$, we shall inductively define the sequences $(\lambda_t^\alpha, n_t, m_t)_{t\geq 1}$, $\alpha \in [0,1]$, and then use the above equations.

6.9. CONSTRUCTION.

Choose and fix sequences $(\delta_t)_{t\geq 1}$, $(p_t)_{t\geq 1}$, and $(\epsilon_t)_{t\geq 1}$ such that

$$\sum_{t=1}^{\infty} \delta_t < 1 \quad , \tag{6.45}$$

$$\begin{cases} p_1 > p_2 > \ldots > 0 \quad , \\ p_t \to 0 \quad \text{as} \quad t \to \infty \quad , \\ \sum_{t=1}^{\infty} p_t = \infty \quad , \end{cases} \tag{6.46}$$

$$\begin{cases} \varepsilon_1 > \varepsilon_2 > \ldots > 0 \ , \\ \sum_{t=1}^{\infty} \varepsilon_t < \infty \ , \end{cases} \tag{6.47}$$

and define

$$\eta_t := \sum_{u=t}^{\infty} \varepsilon_u \ . \tag{6.48}$$

Let λ_1 be a positive number greater than one, and define $\lambda_1^\alpha := \lambda_1$, for all $\alpha \in [0,1]$. Now choose n_1, an integer greater than one, and m_1, an integer greater than $1 + n_1$. Using equations (6.42), (6.43) and (6.44), we now define M_0, N_1, and M_1. Let $A_1 := [0,1]$.

Assume that $(\lambda_u^\alpha, n_u, m_u), m_u > N_u, A_u$, for $u = 1,2,\ldots,t-1$, and $\lambda_1 > \lambda_2 > \ldots > \lambda_{t-1} > 1$, are all chosen.

Choose λ_t such that both

$$1 < \lambda_t < \tfrac{1}{2}(1 + \lambda_{t-1}) \ , \tag{6.49}$$

and, defining $\lambda_t^\alpha := \frac{1}{2}(1 + \lambda_t) + \frac{\alpha}{2}(\lambda_t - 1)$, for all $\alpha \in [0,1]$,

$$\left(\frac{2\lambda_t^\alpha}{1+\lambda_t^\alpha}\right)^{M_{t-1}} < \left(\lambda_t^\alpha\right)^{M_{t-1}} < \exp(\varepsilon_t) \ , \tag{6.50}$$

for all $\alpha \in [0,1]$.

Now take $\rho_t > 0$ satisfying

$$1 < (\lambda_1^\alpha)^{2M_{t-1}} < (\lambda_t^\alpha)^{\rho_t} \ , \tag{6.51}$$

for all $\alpha \in [0,1]$.

Let $c_t > 0$ be such that

$$\frac{1}{\sqrt{2\pi}} \int_{-c_t}^{c_t} \exp(-s^2/2)ds = p_t . \tag{6.52}$$

For any fixed $\alpha \in [0,1]$, we may use the normal distribution's approximation to the binomial distribution to imply that for all large enough integers $n_t^\alpha > M_{t-1}$, both

$$\sum_{U_t^\alpha - 2p_t < |k - \frac{n_t^\alpha \lambda_t^\alpha}{1+\lambda_t^\alpha}| < U_t^\alpha} f_t^\alpha(k) < p_t/4 , \tag{6.53}$$

and

$$1 - 2p_t < \sum_{|k - \frac{n_t^\alpha \lambda_t^\alpha}{1+\lambda_t^\alpha}| > U_t^\alpha} f_t^\alpha(k) \leq 1 - \frac{p_t}{2} , \tag{6.54}$$

where

$$f_t^\alpha(k) = \left(\frac{1}{1+\lambda_t^\alpha}\right)^{n_t-k} \left(\frac{\lambda_t^\alpha}{1+\lambda_t^\alpha}\right)^k \binom{n_t}{k} ,$$

for $k = 0,1,\ldots,n_t$, and

$$U_t^\alpha = (n_t^\alpha)^{\frac{1}{2}}(\lambda_t^\alpha)^{\frac{1}{2}}\ c_t/(1+\lambda_t^\alpha)\quad .$$

Since pointwise convergence implies convergence in probability, we can find an integer $n_t > M_{t-1}$, and a set $A_t \subset [0,1]$ with Lebesgue measure greater than $1 - \delta_t/2$, such that

$$n_t \cdot \left(\frac{\lambda_t - 1}{2(t+1)}\right)^2 > 1\ , \tag{6.55}$$

and equations (6.53) and (6.54) are satisfied for all $\alpha \in A_t$, with n_t^α replaced by n_t . Equations (6.53) and (6.54) imply that

$$\sum_{\left|k - \frac{n_t\lambda_t^\alpha}{1+\lambda_t^\alpha}\right| < U_t^\alpha - 2\rho_t} f_t^\alpha(k) > \rho_t/4\quad , \tag{6.56}$$

for all $\alpha \in A_t$.

For $x \in X$, we define

$$F_t^\alpha(x) := \prod_{u=1}^{t} \left(\frac{2}{1+\lambda_u^\alpha}\right)^{n_u} (\lambda_u^\alpha)^{s_u(x)}\quad , \text{ for } \alpha \in [0,1]\ , \tag{6.57}$$

where $s_u(x) = x_{N_t-N_u+1} + x_{N_t-N_u+2} + \ldots + x_{N_t-M_{u-1}}$,

for $u = 1,2,\ldots,t$, and for $R < S$ (to be specified later)

$$H(x) = \chi_{[R,S]}(x_1+x_2+\ldots+x_{n_t})\ . \tag{6.58}$$

We now use the same arguments as in T. Hamachi's paper [Ha] to show that if q is the measure defined by

$$q := \prod_{i=-\infty}^{\infty} \mu \quad , \tag{6.59}$$

and if E_q denotes integration with respect to q, then the following equations hold for all $\alpha \in A_t$:

$$\lim_{m\to\infty} \frac{\sum_{j=0}^{m} F_t^\alpha(T^j x)H(T^j x)}{\sum_{j=0}^{m} F_t^\alpha(T^j x)} = \frac{E_q(F_t^\alpha H)}{E_q(F_t^\alpha)} \quad , \tag{6.60}$$

$$E_q(F_t^\alpha) = 1 \quad , \tag{6.61}$$

$$E_q(F_t^\alpha H) = \sum_{R\le k\le S} f_t^\alpha(k) \quad , \tag{6.62}$$

$$E_q(F_t^\alpha H_t^\alpha) < 2p_t \quad , \tag{6.63}$$

where H_t^α is defined for $x \in X$ by

$$H_t^\alpha(x) := \chi_{D(t,\alpha)}(x_1 + x_2 + \ldots + x_{n_t}) \quad , \tag{6.64}$$

and $D(t,\alpha)$ is the set given by

$$D(t,\alpha) := \left[\frac{n_t\lambda_t^\alpha}{1+\lambda_t^\alpha} - U_t^\alpha \, , \, \frac{n_t\lambda_t^\alpha}{1+\lambda_t^\alpha} + U_t^\alpha\right] . \tag{6.65}$$

If L denotes Lebesgue measure on the interval $[0,1]$, then since pointwise convergence implies convergence in probability, equations (6.60), (6.61), and (6.63) imply that there exists an integer $J = J(t)$ such that for all $m \geq J$

$$L(\{\alpha\epsilon[0,1];q(\{x\epsilon X;\frac{\sum_{j=0}^{m-1} F_t^{\alpha}(T^j x)H_t^{\alpha}(T^j x)}{\sum_{j=0}^{m-1} F_t^{\alpha}(T^j x)} < 2p_t\}) > 1 - \epsilon_t\}) > 1 - \frac{\delta_t}{2}. \quad (6.66)$$

Choose an integer $m_t > J(t)$ such that both

$$m_t - N_t > J(T) \quad (6.67)$$

and

$$\frac{N_t \exp(2n_{t+1})(\lambda_1)^{3N_t}}{m_t - N_t} < \epsilon_t/2 . \quad (6.68)$$

For $\alpha \epsilon [0,1]$, define the set

$$B_t^{\alpha} := \{x\epsilon X ; \frac{\sum_{j=0}^{m_t-N_t-1} F_t^{\alpha}(T^j x)H_t^{\alpha}(T^j x)}{\sum_{j=0}^{m_t-N_t-1} F_t^{\alpha}(T^j x)} < 2p_t\}$$

and let

$$A_t' := \{\alpha \epsilon A_t ; q(B_t^{\alpha}) > 1 - \epsilon_t\} ,$$

then by equation (6.66) and the choice of the set A_t ,

$$L(A'_t) > 1 - \delta_t \, .$$

We have now chosen λ_t , λ_t^α for $\alpha \in [0,1]$, n_t and m_t , and thus completed step t of the inductive construction.

Let us define the set A by

$$A := \bigcap_{t=1}^{\infty} A'_t \quad ,$$

then by the choice of the sequence $(\delta_t)_{t \geq 1}$, the set A has positive Lebesgue measure, and is thus uncountable. For each $\alpha \in A$, we have an associated sequence $(\lambda_t^\alpha, n_t, m_t)_{t \geq 1}$, that defines a measure

$$p^\alpha := \prod_{k=-\infty}^{\infty} p_k^\alpha \quad ,$$

via the equations (6.40), (6.41), (6.42), (6.43), and (6.44). This measure p^α possesses all the properties used in T. Hamachi's calculations (see [Ha]).

The construction is now complete.

□

6.10. <u>THEOREM</u>. (Ref. [Ha;Theorem 2])

Let T be the shift on the product space X , then with respect to any of the uncountable infinity of measures p^α , for $\alpha \in A$, T is non-singular, ergodic, and admits no σ-finite invariant measure equivalent to p^α . For $\alpha, \beta \in A$, $\alpha \neq \beta$, the measures p^α and p^β are mutually singular.

Proof.

The first part of the theorem is proved by noting that by construction, if $\alpha \in A$ then we can immediately use T. Hamachi's calculations for the measure p^{α}. To prove that the measures are all mutually singular, suppose $\alpha, \beta \in A$ and $\alpha \neq \beta$. By applying Remark 6.5, the following calculation implies that $p^{\alpha} \perp p^{\beta}$.

$$\sum_{i=-\infty}^{\infty} \left(\sum_{j=1}^{2} (p_i^{\alpha}(j) - p_i^{\beta}(j))^2 \right)$$

$$\geq \sum_{t=2}^{\infty} \left(\frac{1}{1+\lambda_t^{\alpha}} - \frac{1}{1+\lambda_t^{\beta}} \right)^2 . n_t \quad ; \text{ by equations (6.41) and (6.40)}$$

$$\geq \sum_{t=2}^{\infty} \left(\frac{\lambda_t^{\beta} - \lambda_t^{\alpha}}{(1+\lambda_1)^2} \right)^2 . n_t$$

$$= \infty \quad ; \text{ using inequality (6.55).}$$

The proof of Theorem 6.10 is complete.

□

6.11. Remark.

Let p^{α}, $\alpha \in A \subset [0,1]$ be a probability measure obtained from Construction 6.9, and let A denote the past σ-algebra corresponding to the shift T, that is the product σ-algebra generated by the co-ordinate maps X_i, for $i \geq 0$. If C denotes the σ-algebra generated by the co-ordinate maps X_i, for $i < 0$, then we can calculate the information cocycle with respect to both A and C (using the measure p^{α}) as in Section 5. By doing this one obtains:

$$J_{p^\alpha}(A,T)(x) = -\log p_0^\alpha(x_0) = \log 2 \ ,$$

and

$$J_{p^\alpha}(C,T)(x) = -\log 2 - \log \frac{dp^\alpha}{dp^\alpha T}(x) \ ,$$

for $x = (x_n) \in X$.

Since there exists no σ-finite T invariant measure equivalent to p^α , and $-\log (dp^\alpha/dp^\alpha T)(x)$ is a recurrent cocycle (because T is conservative with respect to p^α), the cocycle $J_{p^\alpha}(C,T)$ is not cohomologous to a constant.

The two cocycles $J_{p^\alpha}(A,T)$ and $J_{p^\alpha}(C,T)$ are thus fundamentally different; one is constant, and the other cannot even be cohomologous to a constant.

APPENDIX.

ON A BERNOULLI SHIFT WITH NON-IDENTICAL FACTOR MEASURES.

by T. Hamachi.

Ergod. Th. & Dynam. Sys. (1981), 1, 273-283
Printed in Great Britain

On a Bernoulli shift with non-identical factor measures

TOSHIHIRO HAMACHI
Department of Mathematics, Kyushu University, Ropponmatsu, Chuo-Ku, Fukuoka 810, Japan

(*Received* 4 *June* 1981)

Abstract. There exists a Bernoulli shift with non-identical factor measures for which no invariant σ-finite equivalent measure exists.

1. *Introduction*

Our purpose is to give an example of a Bernoulli shift T acting on an infinite product measure space

$$(\Omega, \mathcal{F}, P) = \Big(\prod_{-\infty<k<\infty} \{0,1\}, \bigvee_{-\infty<k<\infty} \mathcal{F}_k, \prod_{-\infty<k<\infty} P_k \Big)$$

such that the shift T,

$$(T\omega)_k = \omega_{k+1} \quad \text{for } \omega = (\omega_k)_{-\infty<k<\infty},$$

is non-singular, i.e. $P(A) = 0$ if and only if $P(TA) = 0$ for $A \in \mathcal{F} = \bigvee_{-\infty<k<\infty} \mathcal{F}_k$ such that T is ergodic and there exists no shift-invariant σ-finite measure equivalent to the infinite product measure $P = \prod_{-\infty<k<\infty} P_k$, where P_k is a probability measure on the set $\{0, 1\}$ and $\mathcal{F}_k$ is the smallest σ-algebra which makes the kth coordinate ω_k of $\omega \in \Omega$ measurable.

This problem was raised by U. Krengel at the symposium on Ergodic Theory at Oberwolfach 1978. He gave in [3] an ergodic Bernoulli shift without finite invariant measure.

In § 2 we give non-ergodic Bernoulli shifts which are dissipative. In § 3 we give an ergodic Bernoulli shift without σ-finite invariant measure.

2. *Dissipative Bernoulli shifts*

Let $\Omega = \prod_{-\infty<k<\infty} \{0, 1\}$, and T be the shift on Ω, i.e.

$$(T\omega)_k = \omega_{k+1}.$$

Take a probability measure

$$P = \prod_{-\infty<k<\infty} P_k$$

with

$$P_k(0) = P_k(1) = \tfrac{1}{2} \quad (k \geq 0)$$
$$0 < P_k(0) < 1, \quad P_k(1) = 1 - P_k(0) \quad (k < 0).$$

It follows from Kakutani's theorem on the equivalence of infinite product measures [2] that T is non-singular if and only if

$$\sum_{-\infty<k<\infty} P_{k-1}(0)P_k(0)((P_{k-1}(1)/P_{k-1}(0))^{\frac{1}{2}} - (P_k(1)/P_k(0))^{\frac{1}{2}})^2 < \infty. \tag{1}$$

In this case we have that the Radon–Nikodym derivative dPT/dP of the measure PT,

$$(PT)(A) = P(TA) \quad \text{for } A \in \mathcal{F},$$

with respect to the measure P is given by

$$(dPT/dP)(\omega) = \prod_{-\infty<k<\infty} P_{k-1}(\omega_k)/P_k(\omega_k) \tag{2}$$

for a.e. ω, where the infinite product converges almost everywhere. In [3] Krengel claimed that the shift T is dissipative if

$$P_k(1)/P_k(0) = 3 \quad (k<0).$$

In fact, more generally we have:

THEOREM 1. *Let* $P_k(0) = P_k(1) = \frac{1}{2}$ $(k \geq 0)$ *and let* $P_k(1)/P_k(0) = \lambda$ (λ *is a constant* >0) $(k<0)$. *If* $\lambda \neq 1$ *then the shift* T *on* $\Omega = \prod_{-\infty<k<\infty}\{0, 1\}$ *is dissipative.*

Proof. The non-singularity condition (1) for the shift T,

$$\begin{aligned}
&\sum_{-\infty<k<\infty} P_{k-1}(0)P_k(0)((P_{k-1}(1)/P_{k-1}(0))^{\frac{1}{2}} - (P_k(1)/P_k(0))^{\frac{1}{2}})^2 \\
&\quad = P_{-1}(0)P_0(0)((P_{-1}(1)/P_{-1}(0))^{\frac{1}{2}} - (P_0(1)/P_0(0))^{\frac{1}{2}})^2 \\
&\quad = (\lambda^{\frac{1}{2}} - 1)^2/(2(1+\lambda)) \\
&\quad < \infty
\end{aligned}$$

is satisfied, and we have from (2)

$$(dPT^n/dP)(\omega) = (2/(1+\lambda))^n \lambda^{S_n(\omega)} \quad (n \geq 1),$$

where $S_n(\omega) = \omega_0 + \omega_1 + \cdots + \omega_{n-1}$.

What we are going to prove is that the infinite series

$$\sum_{n=1}^{\infty} (2/(1+\lambda))^n \lambda^{S_n(\omega)}$$

converges a.e. ω. Take $0<\theta<\frac{1}{2}$, then a standard fact says that

$$\lim_{n\to\infty} (S_n(\omega) - \tfrac{1}{2}n)/(n/4)^{\frac{1}{2}+\theta} = 0 \quad \text{a.e. } \omega.$$

We assume $\lambda > 1$. For any $\varepsilon > 0$ and for a.e. ω, all but a finite number of n satisfy

$$-\varepsilon < (S_n(\omega) - \tfrac{1}{2}n)/(n/4)^{\frac{1}{2}+\theta} < \varepsilon.$$

Then we have for all large n

$$\begin{aligned}
(2/(1+\lambda))^n \lambda^{S_n(\omega)} &< \exp\{n \log(2/(1+\lambda)) + n \log(\lambda)/2 + \varepsilon \log(\lambda)(n/4)^{\frac{1}{2}+\theta}\} \\
&= \exp\{n^{\frac{1}{2}}(\log(2\lambda^{\frac{1}{2}}/(1+\lambda))n^{\frac{1}{2}} + \varepsilon \log(\lambda)(n/4)^{\theta})\}.
\end{aligned}$$

Since for all large n

$$\log(2\lambda^{\frac{1}{2}}/(1+\lambda))n^{\frac{1}{2}} + \varepsilon \log(\lambda)(n/4)^{\theta} < -1,$$

we have for all large n

$$(2/(1+\lambda))^n \lambda^{S_n(\omega)} < \exp(-n^{\frac{1}{2}}).$$

Since the series

$$\sum_{n=1}^{\infty} \exp(-n^{\frac{1}{2}})$$

converges, the theorem is proved if $\lambda > 1$. If $\lambda < 1$, it is enough to see that for a.e. ω and for all large n

$$(2/(1+\lambda))^n \lambda^{S_n(\omega)} < \exp\{n^{\frac{1}{2}}[(\log(2\lambda^{\frac{1}{2}}/(1+\lambda)))n^{\frac{1}{2}} - \varepsilon \log(\lambda)(n/4)^{\varepsilon}]\}. \qquad \square$$

3. *Bernoulli shift without σ-finite invariant measure*

We are concerned with a class of infinite product measures

$$P = \prod_{-\infty<k<\infty} P_k$$

on

$$\Omega = \prod_{-\infty<k<\infty} \{0, 1\}$$

given by

$$P_k = \begin{cases} \mu & \text{if } k \geq 0, \\ \nu_t & \text{if } -N_t < k \leq -M_{t-1}, \\ \mu & \text{if } -M_t < k \leq -N_t \quad (t \geq 1), \end{cases}$$

where

$$N_t = M_{t-1} + n_t, \quad M_t = N_t + m_t, \quad M_0 = 1,$$

n_t and m_t are positive integers, and

$$\mu(0) = \mu(1) = \tfrac{1}{2}, \quad \nu_t(0) = 1/(1+\lambda_t), \quad \nu_t(1) = \lambda_t/(1+\lambda_t),$$

$$\lambda_1 > \lambda_2 > \cdots > \lambda_t > 1 \quad (t \geq 1).$$

We shall also consider the measure

$$Q = \prod_{-\infty<k<\infty} Q_k$$

with $Q_k = \mu$ ($k \in \mathbf{Z}$). The non-singular condition (1) for the shift T on (Ω, P) defined above is equivalent to the condition

$$\sum_{t=1}^{\infty} (\log(\lambda_t))^2 < \infty. \tag{3}$$

What we are going to do is to give inductively a sequence $(\lambda_t, n_t, m_t)_{t\geq1}$ such that the shift T is non-singular, ergodic and admits no invariant σ-finite measures equivalent to P.

3.1. *Construction.* Take sequences $(p_t)_{t\geq1}$ and $(\varepsilon_t)_{t\geq1}$ such that

$$p_1 > p_2 > \cdots > 0, \quad p_t \to 0 \ (\text{as } t \to \infty), \quad \sum_{t=1}^{\infty} p_t = \infty, \tag{4}$$

$$\varepsilon_1 > \varepsilon_2 > \cdots > 0, \quad \sum_{t=1}^{\infty} \varepsilon_t < \infty. \tag{5}$$

and write

$$\eta_t = \sum_{u=t}^{\infty} \varepsilon_u. \tag{6}$$

Let λ_1 be an arbitrary positive number >1, n_1 be an arbitrary integer >1 and m_1 be an arbitrary integer $>1+n_1$. Let

$$M_0 = 1, \quad N_1 = M_0 + n_1, \quad M_1 = N_1 + m_1.$$

We assume that (λ_u, n_u, m_u), $u = 1, 2, \ldots, t-1$ with $\lambda_{u-1} > \lambda_u > 1$, $m_u > N_u$ $(1 \leq u \leq t-1)$ are chosen.

First step: Choice of λ_t. Take λ_t such that

$$1 < \lambda_t < \lambda_{t-1}$$

and

$$(2\lambda_t/(1+\lambda_t))^{M_{t-1}} < \lambda_t^{M_{t-1}} < \exp(\varepsilon_t). \tag{7}$$

Take $\rho_t > 0$ such that

$$1 < (\lambda_1)^{2M_{t-1}} < (\lambda_t)^{\rho_t}. \tag{8}$$

Second step: Choice of n_t. Take $c_t > 0$ such that

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-c_t}^{c_t} \exp(-s^2/2)\, ds = \rho_t.$$

It follows from the central limit theo m that one can obtain a large integer $n_t > M_{t-1}$ such that

$$\sum_{U_t - 2\rho_t \leq |k - n_t\lambda_t/(1+\lambda_t)| \leq U_t} f_t(k) \leq \tfrac{1}{4}\rho_t, \tag{9}$$

$$1 - 2\rho_t < \sum_{|k - n_t\lambda_t/(1+\lambda_t)| > U_t} f_t(k) \leq 1 - \tfrac{1}{2}\rho_t, \tag{10}$$

where

$$f_t(k) = (1/(1+\lambda_t))^{n_t-k}(\lambda_t/(1+\lambda_t))^k \binom{n_t}{k},$$

and

$$U_t = n_t^{\frac{1}{2}}\lambda_t^{\frac{1}{2}}c_t/(1+\lambda_t)$$

for $k = 0, 1, \ldots, n_t$.

By (9) and (10) we have

$$\sum_{|k - n_t\lambda_t/(1+\lambda_t)| < U_t - 2\rho_t} f_t(k) > \tfrac{1}{4}\rho_t. \tag{11}$$

Last step: Choice of m_t. We write

$$F_t(\omega) = \prod_{u=1}^{t} \left(\frac{2}{1+\lambda_u}\right)^{n_u} (\lambda_u)^{\omega_{N_t-N_u+1}+\omega_{N_t-N_u+2}+\cdots+\omega_{N_t-M_{u-1}}}, \tag{12}$$

and for $R < S$

$$H(\omega) = \chi_{(R,S)}(\omega_1 + \omega_2 + \cdots + \omega_{n_t}) \tag{13}$$

for $\omega \in \Omega$. It follows from Birkhoff's ergodic theorem that

$$\lim_{m\to\infty} \frac{\sum_{j=0}^{m} F_t(T^j\omega)H(T^j\omega)}{\sum_{j=0}^{m} F_t(T^j\omega)} = \frac{E_{(t)}(F_tH)}{E_{(t)}(F_t)}. \tag{14}$$

Q-a.e. ω, where E_Q is the integration with respect to the measure Q.

Since

$$(\lambda_u)^{\omega_{N_t-N_u+1}+\omega_{N_t-N_u+2}+\cdots+\omega_{N_t-M_u-1}}, \quad u=1,2,\ldots,t,$$

are independent random variables with respect to the measure Q, we have

$$\begin{aligned} E_Q(F_t) &= \prod_{u=1}^{t}\left(\frac{2}{1+\lambda_u}\right)^{n_u}\int(\lambda_u)^{\omega_{N_t-N_u+1}+\omega_{N_t-N_u+2}+\cdots+\omega_{N_t-M_u-1}}\,dQ(\omega) \qquad (15)\\ &= \prod_{u=1}^{t} 1\\ &= 1, \end{aligned}$$

and

$$\begin{aligned} E_Q(F_tH) &= \prod_{u=1}^{t-1}\left(\frac{2}{1+\lambda_u}\right)^{n_u}\int(\lambda_u)^{\omega_{N_t-N_u+1}+\omega_{N_t-N_u+2}+\cdots+\omega_{N_t-M_u-1}}\,dQ(\omega) \qquad (16)\\ &\quad\times\left(\frac{2}{1+\lambda_t}\right)^{n_t}\int_{R\le\omega_1+\omega_2+\cdots+\omega_{n_t}\le S}(\lambda_t)^{\omega_1+\omega_2+\cdots+\omega_{n_t}}\,dQ(\omega)\\ &= \sum_{R\le k\le S} f_t(k). \end{aligned}$$

We write

$$H_t(\omega)=\chi_{[n_t\lambda_t/(1+\lambda_t)-U_t,\,n_t\lambda_t/(1+\lambda_t)+U_t]}(\omega_1+\omega_2+\cdots+\omega_{n_t}), \qquad (17)$$

then we have from (10) and (16)

$$E_Q(F_tH_t)<2p_t. \qquad (18)$$

It follows from (14), (15) and (18) that we have for all large integers m

$$Q\left(\omega:\frac{\sum_{j=0}^{m-1}F_t(T^j\omega)H_t(T^j\omega)}{\sum_{j=0}^{m-1}F_t(T^j\omega)}<2p_t\right)>1-\varepsilon_t. \qquad (19)$$

We take a large integer m_t with

$$m_t>N_t,$$

such that for

$$A_t=\left\{\omega\in\Omega:\frac{\sum_{j=0}^{m_t-N_t-1}F_t(T^j\omega)H_t(T^j\omega)}{\sum_{j=0}^{m_t-N_t-1}F_t(T^j\omega)}<2p_t\right\}, \qquad (20)$$

we have

$$Q(A_t)>1-\varepsilon_t \qquad (21)$$

and

$$\frac{N_t\exp(2\eta_{t+1})(\lambda_1)^{3N_t}}{m_t-N_t}<\tfrac{1}{2}\varepsilon_t. \qquad (22)$$

THEOREM 2. *Let T be the shift on the infinite product measure space*

$$(\Omega,P)=\left(\prod_{k=-\infty}^{\infty}\{0,1\},\prod_{k=-\infty}^{\infty}P_k\right)$$

constructed above. Then T is non-singular, ergodic and admits no σ-finite invariant measure equivalent to P.

After some preparation we shall prove this theorem.

3.2. *Radon–Nikodym density* $(dPT/dP)(\omega)$. The sequence $(\lambda_i)_{i\geq 1}$ in § 3.1 satisfies the non-singularity condition (3)

$$\begin{aligned}\sum_{i=1}^{\infty}(\log\lambda_i)^2 &< \sum_{i=1}^{\infty}\varepsilon_i^2 \quad (\text{by } (7))\\ &<\infty \quad (\text{by } (5)).\end{aligned}$$

Thus the shift T constructed in § 3.1 is non-singular.

LEMMA 1. *Let T be the shift in* § 3.1 *and put*

$$K_{t,i}(\omega)=\prod_{u=t+1}^{\infty}(\lambda_u)^{-(\omega_{-N_u+1}+\omega_{-N_u+2}+\cdots+\omega_{-N_u+i})+(\omega_{-M_{u-1}+1}+\omega_{-M_{u-1}+2}+\cdots+\omega_{-M_{u-1}+i})},$$

then we have

$$\frac{dPT^i}{dP}(\omega)=K_{t,i}(\omega)\times\prod_{k=-N_t+1}^{i-1}\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)} \quad \text{if } 0\leq i<N_t, \tag{23}$$

and

$$\begin{aligned}\frac{dPT^i}{dP}(\omega)=K_{t,i}(\omega)\times\prod_{u=1}^{t}\left(\frac{1+\lambda_u}{2}\right)^{n_u}(\lambda_u)^{-(\omega_{-N_u+1}+\omega_{-N_u+2}+\cdots+\omega_{-M_{u-1}})}&\\ \times F_t(T^{i-N_t}\omega) \quad \text{if } N_t\leq i<m_t&\end{aligned} \tag{24}$$

for P-a.e. ω, where $F_t(\omega)$ is the random variable defined in (12).

Proof. For $0<i<N_t$ it follows from (2) that

$$\begin{aligned}\frac{dPT^i}{dP}(\omega)&=\prod_{u=t+1}^{\infty}\left\{\prod_{k=-N_u+1}^{-N_u+i}\left(\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\right)\times\prod_{k=-M_{u-1}+1}^{-M_{u-1}+i}\left(\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\right)\right\}\\ &\quad\times\prod_{k=-N_t+1}^{i-1}\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\\ &=\prod_{u=t+1}^{\infty}\left\{\left(\frac{1+\lambda_u}{2}\right)^{i}\times(\lambda_u)^{-(\omega_{-N_u+1}+\omega_{-N_u+2}+\cdots+\omega_{-N_u+i})}\right.\\ &\quad\left.\times\left(\frac{2}{1+\lambda_u}\right)^{i}\times(\lambda_u)^{(\omega_{-M_{u-1}+1}+\omega_{-M_{u-1}+2}+\cdots+\omega_{-M_{u-1}+i})}\right\}\\ &\quad\times\prod_{k=-N_t+1}^{i-1}\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\end{aligned}$$

for P-a.e. ω.

For $N_t\leq i<m_t$, the second factor of (23) is

$$\begin{aligned}&\prod_{u=1}^{t}\left\{\prod_{k=-N_u+1}^{-M_{u-1}}\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\right\}\times\prod_{u=1}^{t}\left\{\prod_{k=i-N_u+1}^{i-M_{u-1}}\frac{P_{k-i}(\omega_k)}{P_k(\omega_k)}\right\}\\ &=\prod_{u=1}^{t}\left\{\left(\frac{1+\lambda_u}{2}\right)^{n_u}(\lambda_u)^{-(\omega_{-N_u+1}+\omega_{-N_u+2}+\cdots+\omega_{-M_{u-1}})}\right\}\\ &\quad\times\prod_{u=1}^{t}\left\{\left(\frac{2}{1+\lambda_u}\right)^{n_u}(\lambda_u)^{(\omega_{i-N_u+1}+\omega_{i-N_u+2}+\cdots+\omega_{i-M_{u-1}})}\right\}\end{aligned}$$

for P-a.e. ω. □

3.3. *Ratio ergodic theorem.* Krengel proved in [3] that if the shift on an infinite product measure

$$\left(\prod_{k=-\infty}^{\infty}\{0,1\},\ \prod_{k=-\infty}^{\infty}P_k\right)$$

with

$$P_k(0)=P_k(1)=\tfrac{1}{2}\quad \text{for } k>0$$

is non-singular and conservative then it is ergodic. The shift in § 3.1 satisfies the condition. This is because

$$F_t(\omega)>\prod_{u=1}^{t}\left(\frac{2}{1+\lambda_u}\right)^{n_u}$$

for every ω, so for $N_t\le i<m_t$ we have from (24)

$$\begin{aligned}\frac{dPT^i}{dP}(\omega)&\ge\prod_{u=t+1}^{\infty}(\lambda_u)^{-(i+1)}\times\prod_{u=1}^{t}(\lambda_u)^{-n_u} \qquad (25)\\ &\ge\prod_{u=t+1}^{\infty}\lambda_u^{-M_{u-1}}\times\prod_{u=1}^{t}\lambda_1^{-n_u}\\ &\ge\prod_{u=t+1}^{\infty}\exp(-\varepsilon_u)\times(\lambda_1)^{-N_t}\quad(\text{by }(7))\\ &=\exp(-\eta_{t+1})\times(\lambda_1)^{-N_t}.\end{aligned}$$

Then we have for a.e. ω,

$$\begin{aligned}\sum_{i=0}^{\infty}\frac{dPT^i}{dP}(\omega)&\ge\sum_{t=1}^{\infty}\sum_{i=N_t}^{m_t-1}\frac{dPT^i}{dP}(\omega)\\ &\ge\sum_{t=1}^{\infty}(m_t-N_t)\exp(-\eta_{t+1})(\lambda_1)^{-N_t}\\ &\ge\sum_{t=1}^{\infty}\frac{2N_t\exp(\eta_{t+1})(\lambda_1)^{2N_t}}{\varepsilon_t}\quad(\text{by }(22))\\ &=\infty.\end{aligned}$$

It follows from the Chacon–Ornstein ratio ergodic theorem [1] that for any measurable set E with $P(E)>0$,

$$\lim_{n\to\infty}\frac{\sum_{i=0}^{n-1}\frac{dPT^i}{dP}(\omega)\chi_E(T^i\omega)}{\sum_{i=0}^{n-1}\frac{dPT^i}{dP}(\omega)}=P(E) \qquad (26)$$

for P-a.e. ω.

PROPOSITION 1. *Let T be the shift constructed in § 3.1 and, for a measurable set E with $P(E)>0$ and $t>1$, we write*

$$\mathrm{I}_t(\omega)=\frac{\sum_{i=0}^{m_t-1}\frac{dPT^i}{dP}(\omega)\chi_E(T^i\omega)}{\sum_{i=0}^{m_t-1}\frac{dPT^i}{dP}(\omega)},$$

$$\mathrm{II}_l(\omega)=\frac{\displaystyle\sum_{i=N_l}^{m_l-1}\frac{dPT^i}{dP}(\omega)\chi_E(T^i\omega)}{\displaystyle\sum_{i=N_l}^{m_l-1}\frac{dPT^i}{dP}(\omega)},$$

$$\mathrm{III}_l(\omega)=\frac{\sum_{i=N_l}^{m_l-1}F_l(T^{i-N_l}\omega)\chi_E(T^i\omega)}{\sum_{i=N_l}^{m_l-1}F_l(T^{i-N_l}\omega)},$$

$$\mathrm{IV}_l(\omega)=\frac{\sum_{i=N_l}^{m_l-1}F_l(T^{i-N_l}\omega)\{1-H_l(T^{i-N_l}\omega)\}\chi_E(T^i\omega)}{\sum_{i=N_l}^{m_l-1}F_l(T^{i-N_l}\omega)},$$

where $H_l(\omega)$ is the random variable defined in (17). *Then we have*

$$\lim_{l\to\infty}\mathrm{I}_l(\omega)=\lim_{l\to\infty}\mathrm{II}_l(\omega)=\lim_{l\to\infty}\mathrm{III}_l(\omega)=\lim_{l\to\infty}\mathrm{IV}_l(\omega)=P(E)$$

for P-a.e. ω.

Proof. We already mentioned in (26) that the limit of $(\mathrm{I}_l)_{l\geq 1}$ exists for P-a.e. ω, and is equal to the constant $P(E)$. Since

$$\begin{aligned}|\mathrm{I}_l(\omega)-\mathrm{II}_l(\omega)|&\leq 2\times\frac{\displaystyle\sum_{i=0}^{N_l-1}\frac{dPT^i}{dP}(\omega)}{\displaystyle\sum_{i=N_l}^{m_l-1}\frac{dPT^i}{dP}(\omega)}\\ &\leq 2\times\frac{N_l\exp(\eta_{l+1})\lambda_1^{2N_l}}{(m_l-N_l)\exp(-\eta_{l+1})\lambda_1^{-N_l}}\quad\text{(by (23) and (25))}\\ &<\varepsilon_l\quad\text{(by (22))},\end{aligned}$$

we have

$$\lim_{m\to\infty}\mathrm{I}_l(\omega)=\lim_{m\to\infty}\mathrm{II}_l(\omega)$$

for P-a.e. ω.

It follows from (24) that

$$\mathrm{II}_l(\omega)=\frac{\sum_{i=N_l}^{m_l-1}K_{l,i}(\omega)F_l(T^{i-N_l}\omega)\chi_E(T^i\omega)}{\sum_{i=N_l}^{m_l-1}K_{l,i}(\omega)F_l(T^{i-N_l}\omega)}.$$

Since

$$\exp(-\eta_{l+1})<K_{l,i}(\omega)<\exp(\eta_{l+1}),$$

we have

$$\exp(-2\eta_{l+1})<\frac{\mathrm{II}_l(\omega)}{\mathrm{III}_l(\omega)}<\exp(2\eta_{l+1}).$$

Thus

$$\lim_{l\to\infty}\mathrm{II}_l(\omega)=\lim_{l\to\infty}\mathrm{III}_l(\omega)$$

for P-a.e. ω.

Let us reconsider the set A_t defined in (21). Since this set is $\bigvee_{k=1}^{M_t-1} \mathcal{F}_k$-measurable, we have

$$P(A_t) = Q(A_t) > 1 - \varepsilon_t.$$

Since

$$\sum_{t=1}^{\infty} P(A_t^c) \le \sum_{t=1}^{\infty} \varepsilon_t < \infty,$$

by the Borel–Cantelli lemma (in the general case), it holds that P-a.e. ω is in A_t for all but a finite number of t. This means that for P-a.e. ω and for all large numbers t

$$\frac{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)H_t(T^{i-N_t}\omega)\chi_E(T^i\omega)}{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)} \le \frac{\sum_{i=0}^{m_t-N_t-1} F_t(T^i\omega)H_t(T^i\omega)}{\sum_{i=0}^{m_t-N_t-1} F_t(T^i\omega)}$$
$$< 2p_t.$$

Since p_t converges to 0, we have that for P-a.e. ω

$$\lim_{t\to\infty} \mathrm{III}_t(\omega) = \lim_{t\to\infty} \frac{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)H_t(T^{i-N_t}\omega)\chi_E(T^i\omega)}{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)} + \lim_{t\to\infty} \mathrm{IV}_t(\omega)$$
$$= \lim_{t\to\infty} \mathrm{IV}_t(\omega). \qquad \square$$

3.4. *Recurrence*

PROPOSITION 2. *Let T be the shift constructed in § 3.1. For any measurable set E with $P(E) > 0$, there exist for P-a.e. ω an infinite number of t such that $T^i\omega \in E$ for some i, $N_t \le i < m_t$, and that for such i, either*

$$\frac{dPT^i}{dP}(\omega) > (\lambda_1)^{2M_{t-1}} \exp(-\eta_{t+1}),$$

or

$$\frac{dPT^i}{dP}(\omega) < (\lambda_1)^{-2M_{t-1}} \exp(\eta_{t+1})$$

holds.

Proof. We define a set B_t by

$$B_t = \left\{\omega \in \Omega : \frac{n_t\lambda_t}{1+\lambda_t} - U_t + 2\rho_t < \omega_{-N_t+1} + \omega_{-N_t+2} + \cdots + \omega_{-M_{t-1}} < \frac{n_t\lambda_t}{1+\lambda_t} + U_t - 2\rho_t\right\}. \tag{27}$$

By (11) we have

$$P(B_t) > \tfrac{1}{2}p_t.$$

Since the set B_t is $\bigvee_{j=-N_t+1}^{-M_{t-1}} \mathcal{F}_j$-measurable, the sets $B_1, B_2, \ldots$ are independent with respect to the measure P. It follows from the Borel–Cantelli lemma (in the independent case) that for P-a.e. ω there exists an infinite number of t such that $\omega \in B_t$.

On the other hand we have from proposition 1 that

$$\lim_{t\to\infty} \frac{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)\{1-H_t(T^{i-N_t}\omega)\}\chi_E(T^i\omega)}{\sum_{i=N_t}^{m_t-1} F_t(T^{i-N_t}\omega)} = P(E)$$

for P-a.e. ω. Then we have that for P-a.e. ω and for all large numbers t, $T^i\omega \in E$, for some i with $N_t \leq i < m_t$, and that for such i either

$$\omega_{i-N_t+1} + \omega_{i-N_t+2} + \cdots + \omega_{i-M_{t-1}} > \frac{n_t\lambda_t}{1+\lambda_t} + U_t$$

or

$$\omega_{i-N_t+1} + \omega_{i-N_t+2} + \cdots + \omega_{i-M_{t-1}} < \frac{n_t\lambda_t}{1+\lambda_t} - U_t$$

holds.

Combining these results, we see that there exists for P-a.e. ω an infinite number of t such that $T^i\omega \in E$ for some i with $N_t \leq i < m_t$, and that for such i,

$$(\lambda_t)^{(\omega_{i-N_t+1}+\omega_{i-N_t+2}+\cdots+\omega_{i-M_{t-1}})-(\omega_{-N_t+1}+\omega_{-N_t+2}+\cdots+\omega_{-M_{t-1}})} > (\lambda_t)^{2\rho_t} \tag{28}$$

or

$$(\lambda_t)^{(\omega_{i-N_t+1}+\omega_{i-N_t+2}+\cdots+\omega_{i-M_{t-1}})-(\omega_{-N_t+1}+\omega_{-N_t+2}+\cdots+\omega_{-M_{t-1}})} < (\lambda_t)^{-2\rho_t} \tag{29}$$

holds. If (28) holds then it follows from (24) that

$$\begin{aligned}\frac{dPT^i}{dP}(\omega) &> (\lambda_t)^{(\omega_{i-N_t+1}+\omega_{i-N_t+2}+\cdots+\omega_{i-M_{t-1}})-(\omega_{-N_t+1}+\omega_{-N_t+2}+\cdots+\omega_{-M_{t-1}})} \\ &\quad \times (\lambda_1)^{-2M_{t-1}} \exp(-\eta_{t+1}) \\ &> (\lambda_t)^{2\rho_t}(\lambda_1)^{-2M_{t-1}} \exp(-\eta_{t+1}) \\ &> (\lambda_1)^{2M_{t-1}} \exp(-\eta_{t+1}) \quad \text{(by (8))}.\end{aligned}$$

If (29) holds then it follows from (24) that

$$\begin{aligned}\frac{dPT^i}{dP}(\omega) &< (\lambda_t)^{(\omega_{i-N_t+1}+\omega_{i-N_t+2}+\cdots+\omega_{i-M_{t-1}})-(\omega_{-N_t+1}+\omega_{-N_t+2}+\cdots+\omega_{-M_{t-1}})} \\ &\quad \times (\lambda_1)^{2M_{t-1}} \exp(\eta_{t+1}) \\ &< (\lambda_t)^{-2\rho_t}(\lambda_1)^{2M_{t-1}} \exp(\eta_{t+1}) \\ &< (\lambda_1)^{-2M_{t-1}} \exp(-\eta_{t+1}).\end{aligned}$$

□

3.5. *Proof of theorem* 2. We assume that T admits a σ-finite invariant measure equivalent to P. Then there exists a positive measurable function $f(\omega)$ such that

$$\frac{dPT^i}{dP}(\omega) = \frac{f(T^i\omega)}{f(\omega)}$$

for $i \in \mathbf{Z}$ and P-a.e. ω. Take $a > b > 0$ such that

$$P(\omega \in \Omega : b < f(\omega) < a) > 0$$

and set $E = \{\omega \in \Omega : b < f(\omega) < a\}$. Then we have

$$\frac{b}{a} < \frac{dPT^i_E}{dP} < \frac{a}{b}$$

for all $i \in \mathbb{Z}$ and P-a.e. ω, that is, the functions $(dPT^i_E/dP)(\omega)$, $i \in \mathbb{Z}$, have a uniformly positive lower bound and a uniform upper bound, where T_E is the induced transformation on the set E of T. However, this contradicts proposition 2. □

Acknowledgement. The author would like to thank W. Krieger for calling his attention to U. Krengel's paper [3], and M. Osikawa and Y. Ito for valuable discussions. He also would like to thank the Institut für Angewandte Mathematik, Universität Heidelberg, and Le Centre de Physique Théorique, CNRS, Marseille, for their hospitality while part of this work was done. The research was supported in part by the Alexander von Humboldt Stiftung.

REFERENCES

[1] R. V. Chacon & D. S. Ornstein. A general ergodic theorem. *Illinois J. Math.* 4 (1960), 153–160.

[2] S. Kakutani. On equivalence of infinite product measures. *Ann. Math.* 49 (1948), 214–224.

[3] U. Krengel. Transformations without finite invariant measure have finite strong generators. Lecture Notes in Math. no. 160, pp. 135–157. Springer-Verlag: Berlin, 1970.

REFERENCES.

A: Anzai, H.: On an example of a measure preserving transformation which is not conjugate to its inverse. Proc. Jap. Acad. 27 (1951), 517-522.

B: Billingsley, P.: Ergodic theory and information. Wiley, New York, 1965.

B&L: Bowen, R., and Lanford, O.E.: Zeta functions of restrictions of the shift transformation. Proc. Symp. Pure Math., Vol. 14 (A.M.S. 1970), 43-49.

Bo1: Bowen, R.: Periodic points and measures for Axiom A diffeomorphisms. Trans. A.M.S. 154 (1971), 377-397.

Bo2: Bowen, R.: Smooth partitions of Anosov diffeomorphisms are weak Bernoulli. Israel J. Math. 21 (1975), 95-100.

Bo3: Bowen, R.: Topological entropy and Axiom A. Proc. Symp. Pure Math., Vol. 14 (A.M.S. 1970), 23-41.

Bo4: Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Mathematics, 470 (1975). Springer-Verlag, Berlin-Heidelberg-New York.

Bu: Butler, R.: M.Sc. Dissertation, Warwick, 1978.

B&S: Butler, R., and Schmidt, K.: An information cocycle for groups of non-singular transformations. Preprint, University of Warwick, 1982.

C&F: Chacon, R.V., and Friedman, N.: Approximation and invariant measures. Zeit. Wahr. Geb. 3 (1965), 286-295.

Ch: Choksi, J.R.: Extension of a theorem of E. Hopf. Journal Lond. Math. Soc. 36 (1961), 81-88. Corrigendum: 36 (1961), 253.

C&K: Cuntz, J., and Krieger, W.: A class of C^*-algebras and topological Markov chains. Invent. math. 56 (1980), 251-268.

D: Doob, J.L.: Stochastic processes. Wiley, New York, 1953.

Du: Dugundji, J.: Topology. Allyn and Bacon, Boston, 1966.

Dy: Dye, H.A.: On groups of measure preserving transformations I. Amer. J. Math. 81 (1959), 119-159.

F&M: Feldman, J., and Moore, C.C.: Ergodic equivalence relations, cohomology and von Neumann algebras I. Trans. A.M.S. 234 (1977), 289-324.

F: Feller, W.: An introduction to probability theory and its applications. Vols. I and II. Wiley, New York, 1957 and 1966.

Fe: Fellgett, R.: Ph.D. Thesis, Warwick, 1976.

F&P: Fellgett, R., and Parry, W.: Endomorphisms of a Lebesgue space II. Bull. Lond. Math. Soc., 7 (1975), 151-158.

Fr: Friedman, N.: Introduction to ergodic theory. Van Nostrand Reinhold Studies No. 29, New York, 1970.

H1: Halmos, P.R.: Lectures on ergodic theory. Chelsea, New York, 1956.

H2: Halmos, P.R.: Measure theory. Van Nostrand, Princeton, N.J., 1950.

H&N: Halmos, P.R., and von Neumann, J.: Operator methods in classical mechanics II. Annals of Mathematics 43 (1942), 332-350.

Ha: Hamachi, T.: On a Bernoulli shift with non-identical factor measures. Ergod. Th. & Dynam. Sys. 1 (1981), 273-283.

I: Ionescu Tulcea, A.: On the category of certain classes of transformations in ergodic theory. Trans. A.M.S. 114 (1965), 261-279.

JKKM&S: del Junco, A., Keane, M., Kitchens, B., Markus, B., and Swanson, L.: Continuous homomorphisms of Bernoulli shifts. Proceedings of the Special Year in Ergodic Theory and Dynamical Systems, 1979-1980, University of Maryland. Progress in Mathematics 10, pages 91-111, Birkhäuser, Stüttgart, 1981.

J&R: del Junco, A., and Rahe, M.: Finitary codings and weak Bernoulli partitions. Proc. A.M.S. 75 (1979), 259-264.

K: Kakutani, S.: On equivalence of infinite product measures. Ann. of Math., 49 (1948), 214-224.

Kr: Krengel, U.: Transformations without finite invariant measure have finite strong generators. Lecture Notes in Mathematics, 160 (1970), 133-157. Springer-Verlag, Berlin-Heidelberg-New York.

Kri1: Krieger, W.: On non-singular transformations of a measure space I. Z. Wahrscheinlichkeitstheorie verw. Geb. 11 (1969), 98-119.

Kri2: Krieger, W.: On dimension functions and topological Markov chains. Invent. math. 56 (1980), 239-250.

K&S: Krzyżewski, K., and Szlenk, W.: On invariant measures for expanding differentiable mappings. Studia Mathematica 33 (1969), 83-92.

L&Y: Lasota, A., and Yorke, J.A.: On the existence of invariant measures for piecewise monotonic transformations. Trans. A.M.S. 186 (1973), 481-488.

M&E: Martin, N.F.G., and England, J.W.: Mathematical Theory of Entropy, Encyclopedia of Mathematics and its Applications., Vol. 12, Addison-Wesley, Reading, 1981.

N: Neveu, J.: Discrete-Parameter Martingales. North-Holland Publishing Co., Amsterdam, 1975.

P1: Parry, W.: Entropy and generators in ergodic theory. Benjamin, New York, 1969.

P2: Parry, W.: Finitary isomorphisms with finite expected code-lengths. Bull. Lond. Math. Soc. 11 (1979), 170-176.

P3: Parry, W.: The information cocycle and ε-bounded codes. Israel J. Math. 29 (1978), 205-220.

P4: Parry, W.: Topics in ergodic theory. Cambridge University Press, Cambridge, 1981.

P5: Parry, W.: An ergodic theorem of information theory without invariant measure. Proc. Lond. Math. Soc. (3) 13 (1963), 605-612.

P6: Parry, W.: Endomorphisms of a Lebesgue space III. Israel J. Math. 21 (1975), 167-72.

P&S: Parry, W., and Schmidt, K.: Preprint, to appear.

P&T1: Parry, W., and Tuncel, S.: On the stochastic and topological structure of Markov chains. Bull. Lond. Math. Soc. 14 (1982), 16-27.

P&T2: Parry, W., and Tuncel, S.: Classification Problems in Ergodic Theory. Lond. Math. Soc. Lecture Note Series, 67. Cambridge University Press, Cambridge, 1982.

P&T3: Parry, W., and Tuncel, S.: On the classification of Markov chains by finite equivalence. Ergod. Th. & Dynam. Sys. 1 (1981), 303-335.

P&W: Parry, W., and Williams, R.F.: Block coding and a zeta function for finite Markov chains. Proc. Lond. Math. Soc. 35 (1977), 483-495.

Pa: Parthasarathy, K.R.: Probability measures on metric spaces. Academic Press, New York - London, 1967.

Pi: Pinsker, M.S.: Information and information stability of random variables and processes. Izdat. Akad. Nauk. SSSR., 1960 (Russian) = Holden-Day, San Francisco, 1964.

Pu&Sh: Pugh, C., and Shub, M.: Ergodicity of Anosov actions. Inventiones math. 15 (1972), 1-23.

R1: Rohlin, V.A.: Lectures on the entropy theory of measure-preserving transformations. Russian Math. Surveys 22 (1967), 1-52.

P4: Parry, W.: Topics in ergodic theory. Cambridge University Press, Cambridge, 1981.

P5: Parry, W.: An ergodic theorem of information theory without invariant measure. Proc. Lond. Math. Soc. (3) 13 (1963), 605-612.

P6: Parry, W.: Endomorphisms of a Lebesgue space III. Israel J. Math. 21 (1975), 167-72.

P&S: Parry, W., and Schmidt, K.: Preprint, to appear.

P&T1: Parry, W., and Tuncel, S.: On the stochastic and topological structure of Markov chains. Bull. Lond. Math. Soc. 14 (1982), 16-27.

P&T2: Parry, W., and Tuncel, S.: Classification Problems in Ergodic Theory. Lond. Math. Soc. Lecture Note Series, 67. Cambridge University Press, Cambridge, 1982.

P&T3: Parry, W., and Tuncel, S.: On the classification of Markov chains by finite equivalence. Ergod. Th. & Dynam. Sys. 1 (1981), 303-335.

P&W: Parry, W., and Williams, R.F.: Block coding and a zeta function for finite Markov chains. Proc. Lond. Math. Soc. 35 (1977), 483-495.

Pa: Parthasarathy, K.R.: Probability measures on metric spaces. Academic Press, New York - London, 1967.

Pi: Pinsker, M.S.: Information and information stability of random variables and processes. Izdat. Akad. Nauk. SSSR., 1960 (Russian) = Holden-Day, San Francisco, 1964.

Pu&Sh: Pugh, C., and Shub, M.: Ergodicity of Anosov actions. Inventiones math. 15 (1972), 1-23.

R1: Rohlin, V.A.: Lectures on the entropy theory of measure-preserving transformations. Russian Math. Surveys 22 (1967), 1-52.

R2: Rohlin, V.A.: Selected topics from the metric theory of dynamical systems. Uspekhi Mat. Nauk, 4 (1949), 57-128 = A.M.S. Translations (2) 49 (1966), 171-240.

Ru: Ruelle, D.: Thermodynamic Formalism. Encyclopedia of Mathematics and its Applications, Vol. 5, Addison-Wesley, Reading, Mass. (1978).

S: Sacksteder, R.: The measures invariant under an expanding map. Lecture Notes in Mathematics, 392 (1974), 179-194. Springer-Verlag, Berlin-Heidelberg-New York.

Sc: Schmidt, K.: Cocycles of ergodic transformation groups. MacMillan (India), 1977.

Se1: Seneta, E.: Non-negative Matrices. Allen and Unwin, London, 1973.

Se2: Seneta, E.: Non-negative Matrices and Markov chains. 2nd. Ed.. Springer Series in Statistics, Springer-Verlag, New York-Heidelberg-Berlin, 1981.

Sh: Shub, M.: Endomorphisms of compact differentiable manifolds. Amer. J. Math. 91 (1969), 175-199.

T1: Tuncel, S.: Conditional pressure and coding. Israel J. Math. 39 (1981), 101-112.

T2: Tuncel, S.: M.Sc. Dissertation, Warwick, 1979.

W1: Walters, P.: Ergodic Theory - Introductory Lectures. Lecture Notes in Mathematics, 458 (1975). Springer-Verlag, Berlin-Heidelberg-New York.

W2: Walters, P.: Invariant measures and equilibrium states for some mappings which expand distances. Trans. A.M.S. 236 (1978), 121-153.